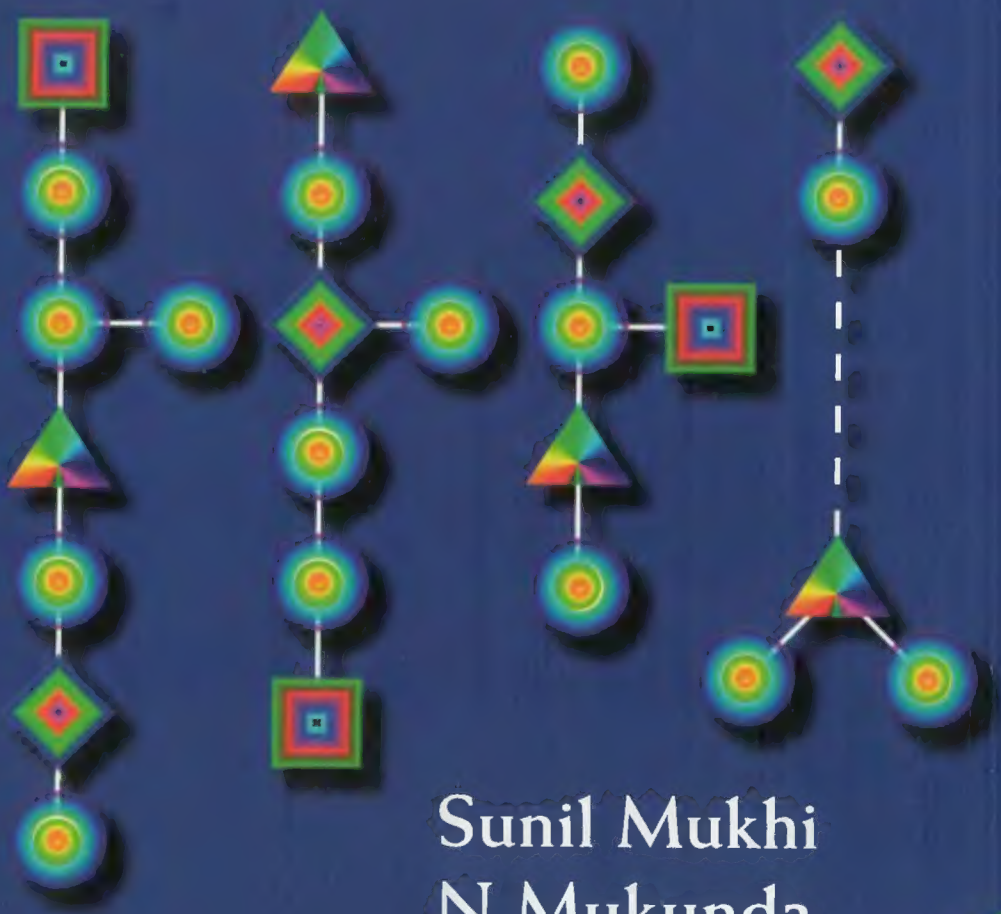
Lectures on **Advanced Mathematical Methods** for **Physicists**



Sunil Mukhi
N Mukunda

HINDUSTAN BOOK AGENCY

World Scientific

Lectures on **Advanced Mathematical Methods** for **Physicists**

Sunil Mukhi
Tata Institute of Fundamental Research, India

N Mukunda
formerly of Indian Institute of Science, India

HINDUSTAN BOOK AGENCY

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

LECTURES ON ADVANCED MATHEMATICAL METHODS FOR PHYSICISTS

ISBN-13 978-981-4299-73-2

ISBN-10 981-4299-73-1

Printed in India, bookbinding made in Singapore.

Contents

Part I: Topology and Differential Geometry

Sunil Mukhi

Department of Theoretical Physics
Tata Institute of Fundamental Research
Mumbai 400 005, India

[+]J.C. Bose Fellow.

Introduction to Part I

These notes describe the basic notions of topology and differentiable geometry in a style that is hopefully accessible to students of physics. While in mathematics the proof of a theorem is central to its discussion, physicists tend to think of mathematical formalism mainly as a tool and are often willing to take a theorem on faith. While due care must be taken to ensure that a given theorem is actually a theorem (i.e. that it has been proved), and also that it is applicable to the physics problem at hand, it is not necessary that physics students be able to construct or reproduce the proofs on their own. Of course, proofs not only provide rigour but also frequently serve to highlight the spirit of the result. I have tried here to compensate for this loss by describing the motivation and spirit of each theorem in simple language, and by providing some examples.

The examples provided in these notes are not usually taken from physics, however. I have deliberately tried to avoid the tone of other mathematics-for-physicists textbooks where several physics problems are set up specifically to illustrate each mathematical result. Instead, the attempt has been to highlight the beauty and logical flow of the mathematics itself, starting with an abstract set of points and adding "qualities" like a topology, a differentiable structure, a metric and so on until one has reached all the way to fibre bundles.

Physical applications of the topics discussed here are to be found primarily in the areas of general relativity and string theory. It is my hope that the enterprising student interested in researching these fields will be able to use these notes to penetrate the dense physical and mathematical formalism that envelops (and occasionally conceals!) those profound subjects.

Chapter 1

Topology

1.1 Preliminaries

Topology is basically the study of continuity. A topological space is an abstract structure defined on a set, in such a way that the concept of continuous maps makes sense. In this chapter we will first study this abstract structure and then go on to see, in some examples, why it is a sensible approach to continuous functions.

The physicist is familiar with ideas of continuity in the context of real analysis, so here we will use the real line as a model of a topological space. For a mathematician this is only one example, and by no means a typical one, but for a physicist the real line and its direct products are sufficient to cover essentially all cases of interest.

In subsequent chapters, we will introduce additional structures on a topological space. This will eventually lead us to manifolds and fibre bundles, the main "stuff" of physics.

The following terms are in common use and it will be assumed that the reader is familiar with them: *set, subset, empty set, element, union, intersection, integers, rational numbers, real numbers.* The relevant symbols are illustrated below:

subset:	$\subset$	empty set:	ϕ	element:	$\in$
union:	$\cup$	intersection:	$\cap$		
set of integers:	$\mathbb{Z}$	set of rational numbers:	Q	set of real numbers:	$\mathbb{R}$

We need some additional terminology from basic set theory that may also be known to the reader, but we will explain it nevertheless.

Definition: If $A \subset B$, the *complement* of A in B, called A' is

$$A' = \{\, x \in B \mid x \notin A \,\}$$

For the reader encountering such condensed mathematical statements for the very first time, let us express the above statement in words: A' is the set of all elements x that are in B such that x is not in A. The reader is encouraged to similarly verbalise any mathematical statement that is not immediately comprehensible.

We continue with our terminology.

Definition: The *Cartesian product* of two sets A and B is the set

$$A \otimes B = \{ (a,b) \mid a \in A,\ b \in B \}$$

Thus it is a collection of ordered pairs consisting of one element from each of the original sets.

Example: The Cartesian product $\mathbb{R} \times \mathbb{R}$ is called $\mathbb{R}^2$, the Euclidean plane. One can interate Cartesian products any number of times, for example $\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ is the n-dimensional Euclidean space $\mathbb{R}^n$.

We continue by defining functions and their basic properties.

Definition: A *function* $A \to B$ is a rule which, for each $a \in A$, assigns a unique $b \in B$, called the *image* of a. We write $b = f(a)$.

A function $f : A \to B$ is *surjective (onto)* if every $b \in B$ is the image of *at least* one $a \in A$.

A function $f : A \to B$ is *injective (one-to-one)* if every $b \in B$ is the image of *at most* one $a \in A$.

A function can be surjective, or injective, or neither, or both. If it is both, it is called *bijective*. In this case, every $b \in B$ is the image of *exactly* one $a \in A$.

Armed with these basic notions, we can now introduce the concept of a topological space.

1.2 Topological Spaces

Consider the real line $\mathbb{R}$. We are familiar with two important kinds of subsets of $\mathbb{R}$:

$$\begin{aligned} \text{Open interval}: \ (a,b) &= \{ x \in \mathbb{R} \mid a < x < b \} \\ \text{Closed interval}: \ [a,b] &= \{ x \in \mathbb{R} \mid a \leq x \leq b \} \end{aligned} \tag{1.1}$$

A closed interval contains its end-points, while an open interval does not. Let us generalize this idea slightly.

Definition: $X \subset \mathbb{R}$ is an *open set in* $\mathbb{R}$ if:

$$x \in X \Rightarrow x \in (a,b) \subset X \text{ for some } (a,b)$$

In words, a subset X of $\mathbb{R}$ will be called *open* if every point inside it can be enclosed by an open interval (a, b) lying entirely inside it.

Examples:
(i) Every open interval (a, b) is an open set, since any point inside such an interval can be enclosed in a smaller open interval contained within (a, b).
(ii) $X = \{ x \in \mathbb{R} \mid x > 0 \}$ is an open set. This is not, however, an open interval.
(iii) $\mathbb{R}$ itself is an open set.
(iv) The empty set ϕ is an open set. This may seem tricky at first, but one merely has to check the definition, which goes "every point inside it can be ...". Such a sentence is always *true* if there are *no* points in the set!
(v) A closed interval is *not* an open set. Take $X = [a, b]$. Then $a \in X$ and $b \in X$ cannot be enclosed by any open interval lying entirely in $[a, b]$. All the remaining points in the closed interval can actually be enclosed in an open interval inside $[a, b]$, but the desired property does not hold for *all* points, and therefore the set fails to be open.
(vi) A single point is not an open set. To see this, check the definition.

Next, we define a closed set in $\mathbb{R}$.

Definition: A *closed set* in $\mathbb{R}$ is the complement in $\mathbb{R}$ of any open set.

Examples:
(i) A closed interval $[a, b]$ is a closed set. It is the complement of the open set $X = \{ x \in \mathbb{R} \mid x > b \} \cup \{ x \in \mathbb{R} \mid x < a \}$.
(ii) A single point $\{a\}$ is a closed set. It is the complement in $\mathbb{R}$ of the open set $X = \{ x \in \mathbb{R} \mid x > a \} \cup \{ x \in \mathbb{R} \mid x < a \}$.
(iii) The full set $\mathbb{R}$ and the empty set ϕ are both closed sets, since they are both open sets and are the complements of each other.

We see that a set can be open, or closed, or neither, or both. For example, $[a, b)$ is neither closed or open (this is the set that contains the end-point a but not the end-point b). Since a cannot be enclosed by an open set, $[a, b)$ is not open. In its complement, b cannot be enclosed by any open set. So $[a, b)$ is not closed either. In $\mathbb{R}$, one can check that the only sets which are both open and closed according to our definition are $\mathbb{R}$ and ϕ.

It is important to emphasize that so far, we are only talking about open and closed sets *in the real line* $\mathbb{R}$. The idea is to extract some key properties of these sets and use those to define open sets and closed sets in an abstract setting. Continuing in this direction, we note some interesting properties of open and closed sets in $\mathbb{R}$:
(a) The union of *any number* of open sets in $\mathbb{R}$ is open. This follows from the definition, and should be checked carefully.
(b) The intersection of a *finite number* of open sets in $\mathbb{R}$ is open.

Why did we have to specify a finite number? The answer is that by taking the intersection of an infinite number of open sets in $\mathbb{R}$, we can actually manu-

facture a set which is not open. As an example, consider the following infinite collection of open intervals:

$$\left\{ \left(-\frac{1}{n}, \frac{1}{n}\right) \;\middle|\; n = 1, 2, 3, \ldots \right\} \tag{1.2}$$

As n becomes very large, the sets keep becoming smaller. But the point 0 is contained in each set, whatever the value of n. Hence it is also contained in their intersection. However any other chosen point $a \in \mathbb{R}$ will lie outside the set $\left(-\frac{1}{n}, \frac{1}{n}\right)$ once n becomes larger than $\frac{1}{|a|}$. Thus the infinite intersection over all n contains only the single point 0, which as we have seen is not an open set. This shows that only finite intersections of open sets are guaranteed to be open.

Having discussed open and closed sets on $\mathbb{R}$, we are now in a position to extend this to a more abstract setting, which is that of a topological space. This will allow us to formulate the concepts of continuity, limit points, compactness, connectedness etc. in a context far more general than real analysis.

Definition: A *topological space* is a set S together with a collection U of subsets of S, called *open sets*, satisfying the following conditions:
(i) $\phi \in U$, $S \in U$.
(ii) The union of arbitrarily many $u_i \in U$ is again in U (thus the union of any number of open sets is open).
(iii) The intersection of finitely many subsets $u_i \in U$ is again in U (thus the intersection of finitely many open sets is open).
The pair (S, U) is called a *topological space*. The collection U of subsets of S is called a *topology* on S.

The reader will notice that if the set S happens to coincide with the real line $\mathbb{R}$, then the familiar collection of open sets on $\mathbb{R}$ satisfies the axioms above, and so $\mathbb{R}$ together with its usual collection of open sets provides an example of a topological space. But as we will soon see, we can put different topologies on the same set of points, by choosing a different collection of subsets as open sets. This will lead to more general (and strange!) examples of topological spaces.

Having defined open sets, it is natural to define closed sets as follows:

Definition: In a given topological space, a *closed set* is a subset of S which is the complement of an open set. Thus, if $u \in U$, then $U' \equiv \{ x \in S \mid x \notin U \}$ is closed.

Examples:
(i) Let S be any set whatsoever. Choose U to be the collection of two sets $\{ \phi, S \}$. This is the smallest collection we can take consistent with the axioms! Clearly all the axioms are trivially satisfied. This is called the *trivial* or sometimes *indiscrete* topology on S.
(ii) Let S again be any set. Choose U to be the collection of *all* subsets of S. This is clearly the largest collection we can possibly take. In this topology, all subsets are open. But they are all closed as well (check this!). This is another

trivial example of a topology, called the *discrete* topology, on S. We see, from this example and the one above, that it is possible to define more than one topology on the same set. We can also guess that if the topology has too few or too many open sets, it is liable to be trivial and quite uninteresting.
(iii)] Let S be the real line $\mathbb{R}$. U is the collection of all subsets X of $\mathbb{R}$ such that

$$x \in X \Rightarrow x \in (a, b) \subset X. \tag{1.3}$$

This is our old definition of "open set in $\mathbb{R}$". We realise now that it was not the unique choice of topology, but it was certainly the most natural and familiar. Accordingly, this topology is called the *usual topology* on $\mathbb{R}$.
(iv) S is a finite set, consisting of, say, six elements. We write

$$S = \{a, b, c, d, e, f\} \tag{1.4}$$

Choose

$$U = \{\phi, S, \{a, b\}, \{b\}, \{b, c\}, \{a, b, c\}\} \tag{1.5}$$

This defines a topology. Some of the closed sets (besides ϕ and S) are $\{d, e, f\}$ and $\{a, c, d, e, f\}$. This example shows that we can very well define a topology on a finite set.

Exercise: If we leave out $\{a, b\}$ in U, do we still get a topology? What if we leave out $\{b\}$? What if we add $\{d\}$?

1.3 Metric spaces

A topological space carries no intrinsic notion of metric, or distance between points. We may choose to define this notion on a given topological space if we like. We will find that among other things, introducing a metric helps to generate many more examples of topological spaces.

Definition: A *metric space* is a set S along with a map which assigns a real number ≥ 0 to each pair of points in S, satisfying the conditions below. If $x \in S$, $y \in S$ then $d(x, y)$ should be thought of as the distance between x and y. The conditions on this map are:
(i) $d(x, y) = 0$ if and only if $x = y$.
(ii) $d(x, y) = d(y, x)$
(iii) $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).
The map $d : S \times S \to \mathbb{R}^+$ is called a *metric* on S.

We see from the list of axioms above, that we are generalising to abstract topological spaces the familiar notion of distance on Euclidean space. Later on we will see that Euclidean space is really a "differentiable manifold with a metric", which involves a lot more structure than we have encountered so far. It is useful to keep in mind that a metric can make sense without any of that other structure – all we need is a set, not even a topological space. In fact we

will see in a moment that defining a metric on a set actually helps us define a topological space.

Examples: On $\mathbb{R}$ we may choose the usual metric $d(x,y) = |x-y|$. On $\mathbb{R}^2$ we can similarly choose $d(x,y) = |\vec{x}-\vec{y}|$. The reader should check that the axioms are satisfied.

Exercise: Does $d(x,y) = (x-y)^2$ define a metric on $\mathbb{R}$?

Given any metric on a set S, we can define a particular topology on S, called the *metric topology*.

Definition: With respect to a given metric on a set S, the *open disc* on S at the point $x \in S$ with *radius* $a > 0$ is the set

$$D_x(a) = \{\, y \in S \mid d(x,y) < a \,\}$$

The open disc is just the set of points strictly within a distance a of the chosen point. However, we must remember that this distance is defined with respect to an abstract metric that may not necessarily be the one familiar to us.

Having defined open discs via the metric, we can now define a topology (called the *metric topology*) on S as follows. A subset $X \subset S$ will be called an open set if every $x \in X$ can be enclosed by some open disc contained entirely in X:

$$X \text{ is open if } x \in X \Rightarrow x \in D_x(a) \subset X \tag{1.6}$$

for some a. This gives us our collection of open sets X, and together with the set S, we claim this defines a topological space.

Exercise: Check that this satisfies the axioms for a topological space.

We realise that the usual topology on $\mathbb{R}$ is just the metric topology with $d(x,y) = |x-y|$. The metric topology also gives the usual topology on n-dimensional Euclidean space $\mathbb{R}^n$. The metric there is:

$$d(x^i, y^i) = \sqrt{\sum_{i=1}^{n} (x^i - y^i)^2} \tag{1.7}$$

where $x^i, y^i,\ i = 1, 2, \cdots, n$ are the familiar coordinates of points in $\mathbb{R}^n$.

The open discs are familiar too. For $\mathbb{R}^2$ they are the familiar discs on the plane (not including their boundary), while for $\mathbb{R}^3$ they are the interior of a solid ball about the point. In higher dimensions they are generalisations of this, so we will use the term "open ball" to describe open discs in any $\mathbb{R}^n$.

Note that for a given set of points we can define many inequivalent metrics. On $\mathbb{R}$, for example, we could use the rather unorthodox (for physicists)

definition:

$$d(x,y) = 1, \quad x \neq y$$
$$= 0, \quad x = y.$$

Thus every pair of distinct points is the same (unit) distance apart.

Exercise: Check that this satisfies the axioms for a metric. What do the open discs look like in this metric? Show that the associated metric topology is one that we have already introduced above.

1.4 Basis for a topology

Defining a topological space requires listing all the open sets in the collection U. This can be quite tedious. If there are infinitely many open sets it might be impossible to list all of them. Therefore, for convenience we introduce the concept of a basis.

For this we return to the familiar case – the usual topology on $\mathbb{R}$. Here, the open intervals (a,b) form a distinguished subclass of the open sets. But they are not *all* the open sets (for example the union of several open intervals is an open set, but is not itself an open interval. If this point was not clear then it is time to go back and carefully review the definition of open sets on $\mathbb{R}$!).

The open intervals on $\mathbb{R}$ have the following properties:

(i) The union of all open intervals is the whole set $\mathbb{R}$:

$$\cup (a_i, b_i) = \mathbb{R} \tag{1.8}$$

(ii) The intersection of two open intervals can be expressed as the union of other open intervals. For example, if $a_1 < a_2 < b_1 < b_2$ then

$$(a_1, b_1) \cap (a_2, b_2) = (a_2, b_1) \tag{1.9}$$

(iii) ϕ is an open interval: $(a,a) = \phi$.

These properties can be abstracted to define a basis for an arbitrary topological space.

Definition: In an arbitrary topological space (S,U), any collection of open sets (a subset of the full collection U) satisfying the above three conditions is called a *basis* for the topology.

A basis contains only a preferred collection of open sets that "generates" (via arbitrary unions and finite intersections) the complete collection U. And there can be many different bases for the same topology.

Example: In any metric space, the open discs provide a basis for the metric

topology. The reader should check this statement.

Exercise: In the usual topology on $\mathbb{R}^2$, give some examples of open sets that are not open discs. Also find a basis for the same topology consisting of open sets that are different from open discs.

1.5 Closure

In many familiar cases, we have seen that the distinction between open and closed sets is that the former do not contain their "end points" while the latter do. In trying to make this more precise, we are led to the concept of *closure* of a set.

In $\mathbb{R}^2$, for example, the set $X_1 = \{ x \subset \mathbb{R}^2 \mid \vec{x}^2 < 1 \}$ defines an open set, the open unit disc. It is open because every point in it can be enclosed in a small open disc which lies inside the bigger one (see Fig. 1.1). Points *on* the unit circle $\vec{x}^2 = 1$ are not in the unit open disc, so we don't have to worry about enclosing them.

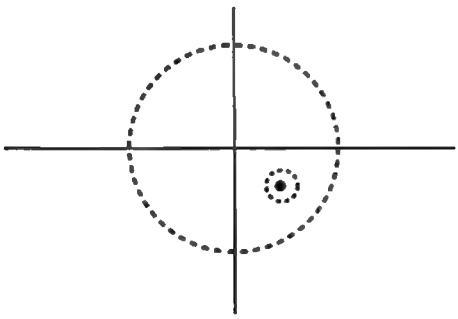


Figure 1.1: The open unit disc. Every point in it can be enclosed in an open disc.

On the other hand consider the set $X_2 = \{ x \subset \mathbb{R}^2 \mid \vec{x}^2 \leq 1 \}$. In addition to points within the unit circle, this set contains all points *on* the unit circle. But clearly X_2 is not an open set, since points on the boundary circle cannot be enclosed by open discs in X_2. In fact X_2 is a *closed* set, as one can check by going back to the axioms.

So it seems that we can add some points to an open set and make it a closed set. Let us make this precise via some definitions.

Definition: Let (S, U) be a topological space. Let $A \subset S$. A point $s \in S$ is called a *limit point* of A if, whenever s is contained in an open set $u \in U$, we have

$$(u - \{s\}) \cap A \neq \phi$$

In other words, S is a limit point of A if every open neighbourhood of S has a non-empty intersection with A.

Exercise: Show that all points on the boundary of an open disc are limit points of the open disc. They do not, however, belong to the open disc. This shows that in general, a limit point of a set A need not be contained in A.

Definition: The *closure* of a set $A \subset S$ is:

$$\bar{A} = A \cup \{ \text{ limit points of } A \}$$

In other words, if we add to a set all its limit points, we get the closure of the set. This is so named because of the following result:

Theorem: The closure of any set $A \subset S$ is a closed set.

Exercise: Prove this theorem. As always, it helps to go back to the definition. What you have to show is that the complement of the closure $\bar{A}$ is an open set.

Given a topological space (S, U), we can define a topology on any subset $A \subset S$. Simply choose U_A to be the collection of sets $u_i \cap A$, $u_i \subset U$. This topology is called the *relative topology* on A. Note that sets which are open in $A \subset S$ in the relative topology need not themselves be open in S.

Exercise: Consider subsets of $\mathbb{R}$ and find an example to illustrate this point.

1.6 Connected and Compact Spaces

Consider the real line $\mathbb{R}$ with the usual topology. If we delete one point, say $\{0\}$, then $\mathbb{R} - \{0\}$ falls into two disjoint pieces. It is natural to say that $\mathbb{R}$ is connected but $\mathbb{R} - \{0\}$ is disconnected. To make this precise and general, we need to give a definition in terms of a given topology on a set.

Definition: A topological space (S, U) is *connected* if it cannot be expressed as the union of two disjoint open sets in its topology. If on the other hand we *can* express it as the union of disjoint open sets, in other words if we can find open sets $u_1, u_2 \in U$ such that $u_1 \cap u_2 = \phi$, $u_1 \cup u_2 = S$, then the space is said to be *disconnected.*

Theorem: In a connected topological space (S, U), the only sets which are both closed and open are S and ϕ.

Exercise: Prove this theorem.

This definition tells us in particular that with the usual topology, $\mathbb{R}^n$ is connected for all n. $\mathbb{R} - \{0\}$ is disconnected, but $\mathbb{R}^n - \{0\}$ is connected for $n \geq 2$. On the other hand, the space $\mathbb{R}^2 - \{ (x, 0) \,|\, x \in \mathbb{R} \}$, which is $\mathbb{R}^2$ with a line removed, is disconnected. Similarly, $\mathbb{R}^2 - \{ (x, y) \,|\, x^2 + y^2 = 1 \}$,

which is $\mathbb{R}^2$ with the unit circle removed, is disconnected. $\mathbb{R}^3$ minus a plane is disconnected, and so on. In these examples it is sufficient to rely on our intuition about connectedness, but in more abstract cases one needs to carefully follow the definition above.

Note that connectivity depends in an essential way on not just the set, but also the collection of open sets U, in other words on the topology. For example, $\mathbb{R}$ with the *discrete* topology is disconnected:

$$\mathbb{R} = \cup_a \{a\}, \quad a \in \mathbb{R} \tag{1.10}$$

Recall that each $\{a\}$ is an open set, disjoint from every other, in the discrete topology. We can also conclude that $\mathbb{R}$ is disconnected in the discrete topology from a theorem stated earlier. In this topology, $\mathbb{R}$ and ϕ are not the only sets which are both closed and open, since each $\{a\}$ is also both closed and open.

We now turn to the study of *closed, bounded* subsets of $\mathbb{R}$ which will turn out to be rather special.

Definition: A *cover* of a set X is a family of sets $\{F_\alpha\} = F$ such that their union contains X, thus

$$X \subset \cup_\alpha F_\alpha$$

If, (S, U) is a topological space and $X \subset S$, then a cover $\{F_\alpha\}$ is said to be an *open cover* if $F_\alpha \in U$ for all α, namely, if F_α are all open sets.

Now there is a famous theorem:

Heine-Borel Theorem: Any open cover of a closed bounded subset of $\mathbb{R}^n$ (in the usual topology) admits a finite subcover.

Let us see what this means for $\mathbb{R}$. An example of a closed bounded subset is an open interval $[a, b]$. The theorem says that if we have *any* (possibly infinite) collection of open sets $\{F_\alpha\}$ which cover $[a, b]$ then a finite subset of this collection also exists which covers $[a, b]$. The reader should try to convince herself of this with an example.

An open interval (a, b) is bounded but not closed, while the semi-infinite interval $[0, \infty) = \{ x \in \mathbb{R} \mid x > 0 \}$ is closed but not bounded. So the Heine-Borel theorem does not apply in these cases. And indeed, here is an example of an open cover of (a, b) with no finite subcover. Take $(a, b) = (-1, 1)$. Let

$$F_n = \left(-1 + \frac{1}{n}, 1 - \frac{1}{n}\right) \quad n = 2, 3, 4, \ldots \tag{1.11}$$

With a little thought, one sees that $\cup_{n=1}^{\infty} F_n = (-1, 1)$. Therefore F_n provides an open cover of $(-1, 1)$. But *no* finite subset of the collection $\{F_n\}$ is a cover of $(-1, 1)$.

Exercise: Find an open cover of $[0, \infty)$ which has no finite subcover.

Closed bounded subsets of $\mathbb{R}^n$ have several good properties. They are

known as *compact* sets on $\mathbb{R}^n$. Now we need to generalize this notion to arbitrary topological spaces. In that case we cannot always give a meaning to "bounded", so we proceed using the equivalent notions provided by the Heine-Borel theorem.

Definition: Given a topological space (S, U), a set $X \subset S$ is said to be *compact* if every open cover admits a finite subcover.

Theorem: Let S be a compact topological space. Then every infinite subset of S has a limit point. This is one example of the special properties of compact sets.

Exercise: Show by an example that this is not true for non-compact sets. It is worth looking up the proof of the above theorem.

Theorem: Every closed subset of a compact space is compact in the relative topology. Thus, compactness is preserved on passing to a topological subspace.

1.7 Continuous Functions

Using the general concept of topological spaces developed above, we now turn to the definition of continuity. Suppose (S, U) and (T, V) are topological spaces, and $f : S \to T$ is a function. Since f is not in general injective (one to one), it does not have an inverse in the sense of a function $f^{-1} : T \to S$. But we can define a more general notion of inverse, which takes any *subset* of T to some *subset* of S. This will provide the key to understanding continuity.

Definition: If $T' \subset T$, then the *inverse* $f^{-1}(T') \subset S$ is defined by:

$$f^{-1}(V) = \{\, s \in S \mid f(s) \in T' \,\}$$

Note that the inverse is defined on all subsets T' of T, including the individual elements $\{t\} \subset T$ treated as special, single-element subsets. However it does not necessarily map the latter to individual elements $s \in S$, but rather to subsets of S as above. The inverse evaluated on a particular subset of T may of course be the empty set $\phi \subset S$.

For the special case of bijective (one-to-one and onto) functions, single-element sets $\{t\} \subset T$ will be mapped to single-element sets $f^{-1}(\{t\}) \subset S$. In this case we can treat the inverse as a function on T, and the definition above coincides with the usual inverse function.

Consider an example in $\mathbb{R}$ with the usual topology:

Example: $f : \mathbb{R} \to \mathbb{R}^+$ is defined by $f(x) = x^2$. Then, $f^{-1} : \{y\} \subset \mathbb{R}^+ \to \{\sqrt{y}, -\sqrt{y}\} \subset \mathbb{R}$ (Fig. 1.2). Now take f^{-1} on an *open set* of $\mathbb{R}^+$, say (1,4). Clearly

$$f^{-1} : (1,4) \subset \mathbb{R}^+ \to \{(1,2), (-1,-2)\} \subset \mathbb{R}. \tag{1.12}$$

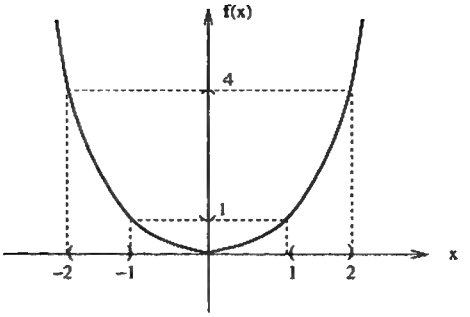


Figure 1.2: $f(x) = x^2$, an example of a continuous function.

Thus the inverse maps open sets of $\mathbb{R}^+$ to open sets of $\mathbb{R}$.

One can convince oneself that this is true for *any* continuous function $\mathbb{R} \to \mathbb{R}^+$. Moreover, it is false for discontinuous functions, as the following example shows (Fig. 1.3):

$$\begin{aligned} f(x) &= x + 1, \quad x < 0 \\ &= x + 2, \quad x \geq 0 \end{aligned}$$

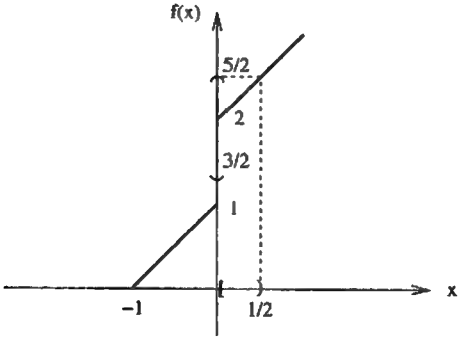


Figure 1.3: An example of a discontinuous function.

In this example, the open set $(\frac{3}{2}, \frac{5}{2}) \subset \mathbb{R}$ is mapped by f^{-1} to the set $[0, \frac{1}{2})$ which is not open. In fact it can be shown that on $\mathbb{R}$ with the usual topology, the continuous functions are those for whom the inverse *always* takes open sets to open sets. For discontinuous functions (i.e. functions having a "break" in their graph) there will be at least one open set which is mapped by the inverse

function to a non-open set.

Exercise: Try to formulate a general proof of the above statement.

As before, we use this property of the real line as a way of defining continuous functions on arbitrary topological spaces.

Definition: For general topological spaces (S, U) and (T, V), a function $f : S \to T$ is called *continuous* if its inverse takes open sets of T to open sets of S.

Exercise: Show that if we put the discrete topology on S and T then *every* function $f : S \to T$ is continuous. Clearly this topology is too crude to capture any interesting information about continuity. Also if we put the indiscrete topology on both S and T then only very few functions are continuous (which are those?). So this topology is also rather uninteresting from the point of view of continuity.

Exercise: Consider 2×2 real matrices

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{1.13}$$

Think of the four entries a, b, c, d as points in $\mathbb{R}^4$ with the metric topology. Show that the "determinant map", $\det : \mathbb{R}^4 \to \mathbb{R}$ is continuous, where $\det(a, b, c, d) = ad - bc$.

1.8 Homeomorphisms

One may define many apparently different topological spaces, but some of these may be "equivalent" for all purposes. To make this notion precise we need to define a kind of map between topological spaces, such that whenever such a map exists, the spaces will be considered to be the same. Basically, what such a map should do is to establish a 1-1 correspondence between elements of the two spaces viewed as sets, in such a way that the open sets of one are mapped onto open sets of the other.

Definition: $f : S \to T$ is a *homeomorphism* if f is bijective and both f and f^{-1} are continuous.

Since f is bijective, we can think of f^{-1} as mapping points of T to unique points of S. The bijective property of f establishes an equivalence between points of S and points of T, while both-ways continuity ensures that open sets go to open sets in each direction.

If a homeomorphism exists between two topological spaces S and T then they are said to be *homeomorphic* and we think of them thereafter as being the same.

Theorem: If $f : S \to T$ is a homeomorphism then S is connected if and only

if T is connected, and S is compact if and only if T is compact.

Exercise: Prove this theorem. This is merely a matter of carefully working through the definitions.

1.9 Separability

We close this chapter with a few definitions and comments concluding with the notion of separability.

Let (S, U) be a topological space.

Definition: The *interior* of a subset $X \subset S$, written X^o, is the union of all open sets u_i contained in X:

$$X^o = \cup_{u_i \subset X, u_i \in U} u_i$$

Definition: The *boundary* of $X \subset S$ is the difference between the closure $\bar{X}$ of X and the interior X^o of X:

$$b(X) = \bar{X} - X^o$$

Examples:
(i) Let $X = (a, b) \subset \mathbb{R}$ in the usual topology. Then $X^o = (a, b)$, $\bar{X} = [a, b]$, and $b(X) = \{a, b\}$. The same is true for each of the choices $X = (a, b]$, $X = [a, b)$, $X = [a, b]$.
(ii) Let $X = \{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1 \}$, the unit open disc. Then $X^o = X$, and $\bar{X} = \{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1 \}$, which is the unit closed disc. The boundary is $b(X) = \{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \}$, which is the unit circle.
(iii) On $\mathbb{R}$ with the usual topology, consider the subset $Q = \{ x \in \mathbb{R} \mid x = \frac{p}{q} \}$ where $p \in \mathbb{Z}, q \in \mathbb{Z} - \{0\}$. These are the rational numbers. One can see that every point of $\mathbb{R}$ is a limit point of Q, because any open interval contains infinitely many rational numbers. Thus $\bar{Q} = \mathbb{R}$. We say that the rationals are *dense* in the reals. Clearly the interior Q^o of Q is ϕ, since no open set fits inside the rationals, so that we also have $b(Q) = \mathbb{R}$.

An important concept in topological spaces is that of *separability*. Given two distinct points, it may be possible to enclose one in an open set which does not contain the other. One may be able to do better, and enclose both of them in two disjoint open sets. There are various degrees of separability which a topological space may have. The one which is most relevant in the study of manifolds and hence of physics is the following:

Definition: A topological space S is *Hausdorff* if, whenever s_1, s_2 are two distinct elements of S, there exist disjoint open sets u_1, u_2 such that $s_1 \in u_1$, $s_2 \in u_2$.

Exercises:
(i) Show that $\mathbb{R}^n$ with the usual topology is Hausdorff.
(ii) Show that $\mathbb{R}$ with the indiscrete topology is not Hausdorff.
(iii) Consider the following type of space. S is any infinite set. $u \subset S$ is an open set if u' is a finite set. Check that this defines a topology on S. Is this a Hausdorff topology?
(iv) Show that every metric space is Hausdorff. In fact, metric spaces are *normal*, a stronger condition implying that disjoint closed sets can be enclosed in disjoint open sets.

Exercise: Show that every compact subset of a Hausdorff space is closed.

Chapter 2

Homotopy

2.1 Loops and Homotopies

In this chapter we discuss ways to understand the connectivity of a topological space. These will consist largely of the study of "closed loops" on a topological space, and the possibility of deforming these into each other. Many, though not all, essential properties of a topological space emerge on studying connectivity.

We have already defined a *connected* topological space: one which cannot be expressed as the union of two disjoint open sets. There is another kind of "connectivity" property of topological spaces which will prove very important. Consider as an example the plane $\mathbb{R}^2$ with the unit disc cut out (Fig. 2.1).

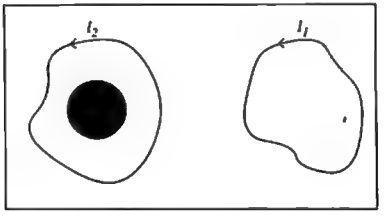


Figure 2.1: $\mathbb{R}^2$ with a disc cut out.

On this space, a loop like l_1 (we will give a precise definition of "loops" later), which does not encircle the disc, has the property that:
(i) l_1 can be continuously shrunk to a point.
(ii) l_1 can be continuously deformed to any other loop not encircling the disc.
(iii) l_1 cannot be continuously deformed to a loop like l_2 which encircles the disc once.

The study of whether loops in a topological space can be deformed into others is part of *homotopy* theory, and is an important tool in characterizing

the topology of spaces. In addition to loops, which are topologically circles, we can consider subspaces that are topologically higher dimensional spheres S^n. For example, $\mathbb{R}^3$ with the unit ball removed has a different connectivity property: all loops can be deformed into each other, but 2-spheres in that space may not be shrinkable, if they enclose the "hole".

So far we have worked at an intuitive level and with familiar spaces. Now let us give precise definitions of the objects in terms of which we will formulate the study of homotopy.

Definition: A *path* $\alpha(t)$ in a topological space S, from $x_0 \in S$ to $x_1 \in S$, is a continuous map

$$\alpha : [0,1] \to S$$

such that

$$\alpha(0) = x_0, \quad \alpha(1) = x_1$$

Note that the space $[0,1]$ appearing on the left-hand-side of the definition is the familiar closed unit interval in $\mathbb{R}$, while the space on the right-hand-side can be any arbitrary topological space.

When S is a space like $\mathbb{R}^n$ with which we are familiar, the above definition of a path seems quite reasonable. But continuous paths can exist in arbitrary topological spaces, even including spaces with finitely many points! On reflection this should be no surprise, since our definition of continuity in the previous chapter made sense for arbitrary topological spaces. A simple example is provided by the following exercise.

Exercise: Consider the set $S = \{a, b, c\}$ of three elements, with the topology $U = (\phi, S, \{a\}, \{b\}, \{a, b\})$. Find a continuous path from a to b.

It should be kept in mind that a path as defined above is not just the *image* of a map, but the map itself. One should imagine that as the parameter t moves through values from 0 to 1, its image in the topological space moves in that space, and it is the map between these two motions which we call the path. It is convenient to think of the parameter t as a *time*, then the map gives the motion of a point on the topological space as a function of time. For example, given a path $\alpha(t)$, we can define a new path $\beta(t) = \alpha(t^2)$ which traces out the same image in the same total time, but corresponds to a different map and is therefore treated as a different path.

Having defined paths, we use them to define a new notion of connectedness.

Definition: S is *arcwise connected* or *path connected* if there always exists a path $\alpha(t)$ between any pair of points x_0 and x_1.

Theorem: If S is arcwise connected then it is connected. (Recall that we defined connectedness in the previous chapter, without recourse to paths in the space.)

Exercise: Prove the above theorem. Assume S is arcwise connected, but not

connected, and find a contradiction.

A path, as defined above, has distinct end-points in general. But paths which close on themselves turn out to be the most useful ones.

Definition: A *closed path* or *loop* in S at x_0 is a path $\alpha(t)$ for which $x_0 = x_1$, that is, $\alpha(0) = \alpha(1) = x_0$. The loop is said to be *based* at x_0 (see Fig. 2.2).

Notice that a loop is a map from a *circle* to an arbitrary topological space S. However it contains some additional information in the form of the "base point" x_0. In what follows, we will always deal with based loops, in other words loops with a fixed base point.

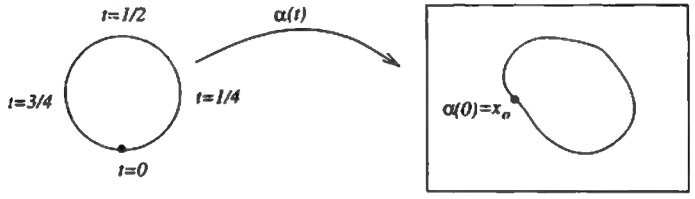


Figure 2.2: A based loop in a topological space S.

We will find it useful to develop the notion of *multiplication of loops*. Each loop is a map from the circle to the space, and the product will again be such a map. We only admit multiplication between two based loops with the same base point.

To multiply two loops based at the same point x_0, define a map $[0,1] \to S$ for which, as t goes from 0 to 1, the image first traces out one loop and then the other. This is formalised as follows.

Definition: Suppose $\alpha(t), \beta(t)$ are two loops based at x_0. The *product loop* $\gamma = \alpha * \beta$ is defined by:

$$\begin{aligned} \gamma(t) &= \alpha(2t) && 0 \le t \le \frac{1}{2} \\ &= \beta(2t-1) && \frac{1}{2} \le t \le 1 \end{aligned} \tag{2.1}$$

It is easy to check that this defines a continuous loop based at x_0. Thinking of t as time, we may say that in the product map defined above, the image point in S moves along the loop α during the first half of the total time, and along β in the second half.

Definition: The *inverse loop* $\alpha^{-1}(t)$ of a loop $\alpha(t)$ is defined by:

$$\alpha^{-1}(t) = \alpha(1-t), \quad 0 \le t \le 1$$

This is just the same loop traced backwards. If we only look at their images in the topological space, a loop and its inverse look the same, but as maps they are clearly different.

Definition: The *constant loop* is the map $\alpha(t) = x_0,\ 0 \leq t \leq 1$. The image of this map is a single point.

Now we implement the intuitive idea that some loops based at x_0 can be continuously deformed to other loops based at the same point. Such a deformation, if it exists, should be a map which specifies how a given loop "evolves" continuously to another one.

Definition: Two loops $\alpha(t)$ and $\beta(t)$ based at x_0 are *homotopic* to each other if there exists a continuous map:

$$H : [0,1] \times [0,1] \to S$$

such that

$$\begin{aligned} H(t,0) &= \alpha(t), & 0 \leq t \leq 1 \\ H(t,1) &= \beta(t), & 0 \leq t \leq 1 \\ H(0,s) &= H(1,s) = x_0, & 0 \leq s \leq 1 \end{aligned}$$

The meaning of this is obvious with a little thought. We have introduced a new parameter $s \in [0,1]$ which we would also like to think of as a second "time". The map $H(t,s)$, for each fixed s, defines a loop in the space S based at x_0. Therefore, as s evolves, the entire loop can be thought of as evolving. At the initial value $s = 0$ the loop was $\alpha(t)$, while at the final value $s = 1$ it has become a different loop $\beta(t)$. At every intermediate value of s it is some other loop, *always based at* x_0. H is called a *homotopy*. An example of a homotopy $H(t,s)$ is illustrated in Fig. 2.3.

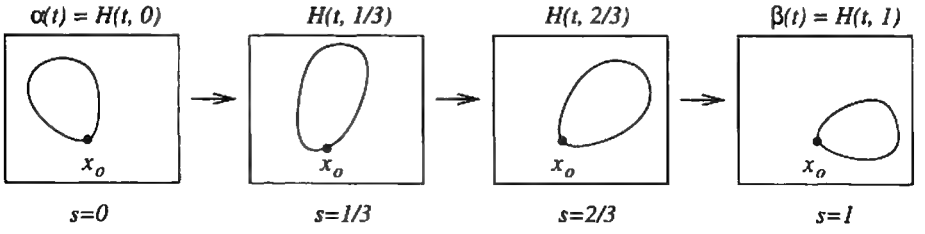


Figure 2.3: A homotopy between two based loops $\alpha(t)$ and $\beta(t)$.

If two curves $\alpha(t), \beta(t)$ are homotopic to each other, we write $\alpha \stackrel{\cdot}{\sim} \beta$.

Theorem: Homotopy is an *equivalence relation*. This means the following conditions are satisfied:

(i) $\alpha \stackrel{\cdot}{\sim} \alpha$ (reflexivity)
(ii) $\beta \stackrel{\cdot}{\sim} \alpha \Leftrightarrow \alpha \stackrel{\cdot}{\sim} \beta$ (symmetry)
(iii) $\alpha \stackrel{\cdot}{\sim} \beta, \beta \stackrel{\cdot}{\sim} \gamma \Rightarrow \alpha \stackrel{\cdot}{\sim} \gamma$ (transitivity)

Proof: Here is a sketch of the proof, which consists of displaying explicitly a homotopy H in each case.
(i) Pick $H(t,s) = \alpha(t)$ for all s. This proves reflexivity.
(ii) Given $H(t,s) : \alpha \to \beta$, define a new homotopy $H(t, 1-s) : \beta \to \alpha$. This proves symmetry.
(iii) Given $H_1(t,s) : \alpha \to \beta$, and $H_2(t,s) : \beta \to \gamma$, define $H_3(t,s) : \alpha \to \gamma$ by

$$\begin{aligned} H_3(t,s) &= H_1(t,2s), \qquad 0 \le s \le \frac{1}{2} \\ &= H_2(t,2s-1), \quad \frac{1}{2} \le s \le 1 \end{aligned} \tag{2.2}$$

This proves transitivity and completes the proof.

Homotopies also satisfy the following additional properties:

(iv) $\alpha \stackrel{\cdot}{\sim} \beta \Leftrightarrow \alpha^{-1} \stackrel{\cdot}{\sim} \beta^{-1}$
(v) $\alpha \stackrel{\cdot}{\sim} \beta, \alpha' \stackrel{\cdot}{\sim} \beta' \Rightarrow \alpha \star \alpha' \stackrel{\cdot}{\sim} \beta \star \beta'$.

However, it is *not* in general true that $\alpha \stackrel{\cdot}{\sim} \alpha^{-1}$.

Exercise: Find an example of a topological space and a path $\alpha(t)$ in it such that α is not homotopic to α^{-1}. Find another example for which α and α^{-1} are nontrivial yet homotopic to each other.

2.2 The Fundamental Group

A very important property of an equivalence relation on a set of objects (in this case the set of based loops) is that it partitions the set into disjoint classes, called "equivalence classes". Within each class, all elements are equivalent to each other under the given relation.

In the present case, denote by $[\alpha]$ the class of all loops homotopic to $\alpha(t)$ (always with reference to a fixed base point). One can check that multiplication of classes

$$[\alpha] \star [\beta] = [\alpha \star \beta]$$

is well-defined. This amounts to showing that the above product defines the same class irrespective of which loops we choose as representatives for the classes on the left hand side.

Definition: The collection of *all distinct* homotopy classes of loops in X based at x_0 is:

$$\pi_1(S, x_0) \equiv \{\ [\alpha] \mid \alpha(t) \text{ is a loop in } S \text{ based at } x_0\ \}$$

$\pi_1(S, x_0)$ is a *group* under multiplication. It is called the *fundamental group* or *first homotopy group* of S at x_0.

The group property of π_1 needs to be proved. We do this by checking the group axioms:

(i) *Closure:*

$$[\alpha] \in \pi_1,\ [\beta] \in \pi_1 \Rightarrow [\alpha] \star [\beta] = [\alpha \star \beta] \in \pi_1$$

(ii) *Identity:* $[i]$ is the equivalence class of loops homotopic to the constant loop at x_0. Clearly

$$[\alpha] \star [i] = [i] \star [\alpha] = [\alpha \star i] = [\alpha]$$

for all $[\alpha]$.

(iii) *Inverse:* $[\alpha]^{-1} = [\alpha^{-1}]$, because

$$[\alpha]^{-1} \star [\alpha] = [\alpha^{-1} \star \alpha] = [i]$$

(iv) *Associativity:* To show that

$$([\alpha] \star [\beta]) \star [\gamma] = [\alpha] \star ([\beta] \star [\gamma])$$

is straightforward but requires some thought, and this part of the proof is left as an exercise to the reader.

To procced further, we will find it useful to define the product operation even on paths which are *not* closed, namely, paths $\alpha(t)$ such that $\alpha(0)$ and $\alpha(1)$ are distinct.

Definition: The *product* of two *open paths* $\alpha(t)$ and $\beta(t)$ is defined only if $\alpha(1) = \beta(0)$, in other words, if the final point of α is the same as the initial point of β, and is given by $\gamma = \alpha \star \beta$ where:

$$\begin{aligned} \gamma(t) &= \alpha(2t), \qquad 0 \le t \le \frac{1}{2} \\ &= \beta(2t-1), \qquad \frac{1}{2} \le t \le 1 \end{aligned} \tag{2.3}$$

As one might expect, the initial point of the product path $\alpha \star \beta$ is the initial point of α, and the final point of $\alpha \star \beta$ is the final point of β.

Now returning to the fundamental group, we note that in principle this group depends on the base point with respect to which it is defined. Thus, $\pi_1(S, x_0)$ is a different group from $\pi_1(S, x)_1$ for $x_0 \neq x_1$. But in fact under very general conditions the two are the same.

To specify what we mean by two groups being the "same", let us first define the concept of "homomorphism" between groups (not be confused with "homeomorphism" between topological spaces!). This is a mapping from one

group to another which preserves the group operations.

Definition: If G and H are two groups, a *homomorphism* $\phi : G \to H$ is a map $g \in G \to \phi(g) \in H$ such that:

$$\phi : g^{-1} \in G \Rightarrow \phi(g^{-1}) = (\phi(g))^{-1} \in H$$
$$\phi : g_1 \cdot g_2 \in G \Rightarrow \phi(g_1 \cdot g_2) = \phi(g_1) \cdot \phi(g_2) \in H$$
$$\phi : i \in G \Rightarrow \phi(i) = i' \in H$$

(here "ϕ :" should be read as "ϕ maps" (a given element of G to an element of H)).

Definition: A homomorphism ϕ is an *isomorphism* if it is also bijective (one-to-one and onto). If an isomorphism exists between two groups, the groups are said to be *isomorphic*, and can be thought of as completely equivalent to each other.

We now show that the fundamental group of a topological space is independent of the base point under some conditions.

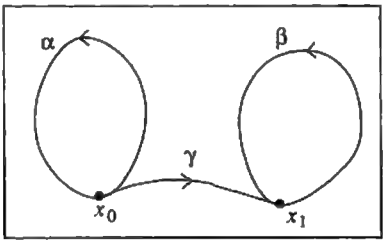


Figure 2.4: Isomorphism of $\phi_1(S, x_0)$ and $\phi_1(S, x_1)$ for a path-connected space.

Theorem: If a topological space S is path-connected, $\pi_1(S, x_0)$ and $\pi_1(S, x_1)$ are isomorphic as groups.

Proof: This is illustrated in Fig. 2.4. Consider $\alpha(t)$ based at x_0 along with its equivalence class $[\alpha] \in \pi_1(S, x_0)$. Let $\gamma(t)$ be an open path from x_0 to x_1. Now define a map

$$\phi : \pi_1(S, x_0) \to \pi_1(S, x_1) \tag{2.4}$$

by

$$\phi([\alpha]) = [\gamma^{-1} \star \alpha \star \gamma] \tag{2.5}$$

Clearly $[\alpha] \in \pi_1(S, x_0) \Rightarrow \phi([\alpha]) \in \pi_1(S, x_1)$. And moreover this is an isomor-

phism. For example, the product law holds:

$$\begin{aligned}\phi([\alpha]\star[\beta]) &= \phi([\alpha\star\beta]) \\ &= [\gamma^{-1}\star\alpha\star\beta\star\gamma] \\ &= [\gamma^{-1}\star\alpha\star\gamma]\star[\gamma^{-1}\star\beta\star\gamma] \\ &= \phi[(\alpha)]\star\phi([\beta])\end{aligned} \tag{2.6}$$

with $[\alpha]$, $[\beta] \in \pi_1(S, x_0)$. Similarly one can check the other properties of a homomorphism, as well as the fact that this map is bijective which finally makes it an isomorphism.

Because of the above theorem, for path-connected spaces we denote the fundamental group as $\pi_1(S)$ instead of $\pi_1(S, x_0)$. It must be kept in mind, however, that in the process of finding π_1 we must always work with loops based at some (arbitrary) point x_0. The final result for π_1 will then be independent of x_0.

2.3 Homotopy Type and Contractibility

So far we have discussed maps from the closed interval [0, 1] to a topological space S, and defined two such maps to be homotopic if there exists a suitable map from $[0,1]\times[0,1]$ to S. This has a straightforward generalisation involving two topological spaces.

Definition: Let S and T be two arbitrary topological spaces, and consider two different continuous maps

$$\begin{aligned} f_0 &: S \to T \\ f_1 &: S \to T \end{aligned} \tag{2.7}$$

The maps f_0 and f_1 are said to be *homotopic* to each other if there exists a continuous map

$$F : S \otimes [0,1] \to T \tag{2.8}$$

such that

$$\begin{aligned} F(x,0) &= f_0(x) \\ F(x,1) &= f_1(x) \end{aligned} \tag{2.9}$$

In the special case where S is the closed interval $[0,1]$ on the real line, this homotopy of maps reduces to homotopy of paths or loops, but one should note that there is no reference to a base-point. So we will use the symbol $\sim$ to denote the homotopy of maps defined above (which does not involve a base point), as against $\dot{\sim}$ which always denotes the homotopy of based loops.

Homotopy of maps between general topological spaces is useful because it helps us identify properties that the two spaces may have in common. For

example, let us find the condition that two different path-connected spaces S and T have the same fundamental group: $\pi_1(S) = \pi_1(T)$.

Definition: Two topological spaces S and T are of the *same homotopy type* if there exist continuous maps

$$f: S \to T \qquad \text{and} \qquad g: T \to S$$

such that (we use the symbol "o" for composition of maps):

$$f \circ g: T \to T \sim i_T$$
$$g \circ f: S \to S \sim i_S$$

where i_T is the identity map $T \to T$ and similarly for i_S.

The property of "being of the same homotopy type" guarantees that $\pi_1(S)$ and $\pi_1(T)$ are the same group. This is embodied in the following result.

Theorem: If S and T are two path-connected topological spaces of the same homotopy type, then $\pi_1(S)$ is isomorphic as a group to $\pi_1(T)$. (The proof is somewhat complicated and we will skip it.)

Theorem: If S and T are homeomorphic as a topological spaces then in particular they are of the same homotopy type, and hence have the same π_1. This is obvious from the above, since a homeomorphism is a pair of continuous maps $f: S \to T$, $g: T \to S$ such that $g = f^{-1}$, i.e.

$$f \circ g = i_T, \qquad g \circ f = i_S. \tag{2.10}$$

Thus the theorem is true.

Summarizing, we have found out two important facts:

(i) Homotopy is a *topological invariant.* Two spaces which are topologically equivalent (homeomorphic) have the same homotopy properties.
(ii) However, the converse is not necessarily true. Two topological spaces may have the same π_1 (if they are of the same homotopy type), but this does *not* imply that they are homeomorphic.

The following definitions, theorems and examples tend to crop up fairly often in physical applications.

Definition: A topological space is *contractible* if it is of the same homotopy type as a single point.
Definition: A topological space S for which $\pi_1(S) = \{i\}$ is called *simply connected.* Otherwise it is called *multiply connected.*

A simply connected space has no "nontrivial" loops, in other words all loops are deformable, or homotopic, to each other.

For a contractible space S, the fundamental group $\pi_1(S) = \{i\}$, the identity element. Therefore it is simply connected. However a simply connected space need not be contractible, as we will see in examples below.

Theorem: $\pi_1(x \otimes y) = \pi_1(x) \oplus \pi_1(y)$. Here the direct sum of groups, $\oplus$, is their Cartesian product as sets:

$$\pi_1(x) \oplus \pi_1(y) = \{\ (a, b)\ |\ a \in \pi_1(x),\ b \in \pi_1(y)\ \} \tag{2.11}$$

The proof of this theorem is quite simple and is left as an exercise.

Examples:

(i) $\mathbb{R}^n$ is contractible. To show this, we need to find continuous maps $f : \mathbb{R}^n \to \{p\}$ and $g : \{p\} \to \mathbb{R}^n$. Clearly the only available choices are $f(\vec{x}) = p$, the constant map, and $f(p) = \vec{0}$, where 0 is some chosen point in $\mathbb{R}^n$ (which we may call the origin). Then, $g \circ f : \mathbb{R}^n \to \mathbb{R}^n$ is the map

$$g \circ f\ (\vec{x}) = \vec{0} \tag{2.12}$$

Define a homotopy $F : \mathbb{R}^n \times [0, 1] \to \mathbb{R}^n$ by $F(\vec{x}, t) = t\vec{x}$. For $t = 0$, this is the constant map $g \circ f$ which sends all points to 0. For $t = 1$, it is the identity map $i_{\mathbb{R}^n}$. Thus $g \circ f \sim i_{\mathbb{R}^n}$. Of course $f \circ g$ is trivially $\sim i$. Thus we have shown that $\mathbb{R}^n$ is of the same homotopy type as a point, and hence contractible. It follows that $\pi_1(\mathbb{R}^n) = \{i\}$. (This is also intuitively obvious, for any loop in $\mathbb{R}^n$ is homotopically trivial.)

(ii) S^2 is not contractible (this is illustrated in Fig. 2.5). To prove contractibility of a space, we must find a way to "move" all its points continuously to one point. On S^2 we could try to move everything to the south pole along great circles, but then the north pole cannot move in any direction without breaking continuity! Nevertheless, any based loop on S^2 can be continuously deformed to a point, therefore $\pi_1(S^2) = \{i\}$ and S^2 is simply connected.

We see that from the point of view of contractibility S^2 is nontrivial, but from the point of view of the fundamental group it is trivial.

(iii) S^1, the circle, is not contractible. The proof is just as in the previous example. But it also has a nontrivial fundamental group, unlike the previous case. To show this, let $\alpha_n(t)$ be a loop which winds n times around the circle in an anticlockwise direction for $n > 0$, and $|n|$ times clockwise for $n < 0$. For $n = 0$, take the constant loop, namely, a point $x_0 \in S^1$. Clearly $[\alpha_n] \neq [\alpha_m]$ for $m \neq n$, and the collection $[\alpha_n]$, $n \in \mathbb{Z}$ exhausts all homotopy classes. Under multiplication,

$$[\alpha_n] \star [\alpha_m] = [\alpha_{n+m}] \tag{2.13}$$

Thus, $\pi_1(S^1) = \{\ [\alpha_n],\ n \in \mathbb{Z}\ \}$, with $[\alpha_n] \star [\alpha_m] = [\alpha_{n+m}]$. Clearly this group is isomorphic to the integers under addition. The isomorphism is:

$$[\alpha_n] \in \pi_1(S^1)\ \to\ n \in \mathbb{Z} \tag{2.14}$$

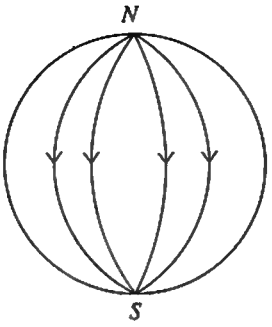


Figure 2.5: S^2 is not contractible: we can move all points except the north pole to the south pole.

The isomorphism permits us to write $\pi_1(S^1) = \mathbb{Z}$. The number n labelling the class to which a given loop $\alpha(t)$ belongs is called the *winding number* of the loop.

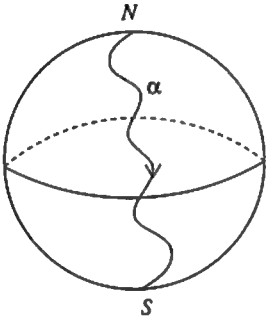


Figure 2.6: A closed loop in S^2/θ.

(iv) Consider the two-sphere S^2 with diametrically opposite points identified. This is the first seriously nontrivial space we are studying! If the identification map is called θ then we may think of the space as the quotient S^2/θ. A representative set of points is the upper hemisphere, which is identified with the lower hemisphere, as well as half the equator, which is identified with the other half.

Now all loops which were closed in S^2 will remain closed in S^2/θ and of course they are still shrinkable. However, there are paths which are open in S^2 but closed in S^2/θ. For this, consider any path that starts at the north pole and ends at the south pole, as is illustrated in Fig. 2.6. This path is closed in S^2/θ because the north and south poles are identified, but it is not shrinkable to a point.

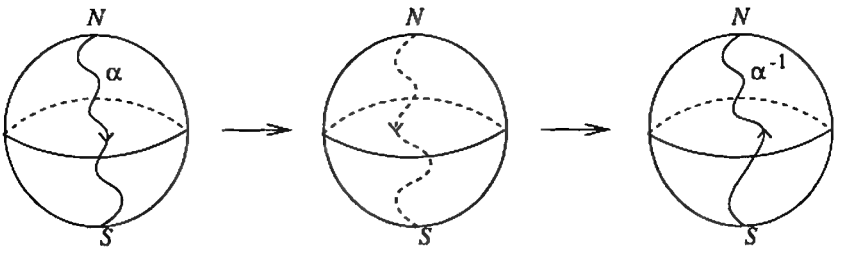


Figure 2.7: α is deformable to α^{-1} on S^2/θ.

Now let us show that this loop α is homotopic to its inverse α^{-1} (see Fig. 2.7). Simply keep the end points fixed and deform the path so that it goes down the other side of the sphere. By the identification map, this is the same as a path going *upwards* from S to N along the front of the sphere, namely, α^{-1}. Thus we have shown that in this case,

$$[\alpha] = [\alpha^{-1}] = [\alpha]^{-1} \tag{2.15}$$

So

$$[\alpha] \star [\alpha] = [\alpha^2] = [\alpha] \star [\alpha]^{-1} = [i] \tag{2.16}$$

It is clear that a shrinkable loop based at N (which is closed already in S^2) and the loop $\alpha(t)$ we have just described (which is closed only on S^2/θ) represent all the homotopy classes of S^2/θ. Thus

$$\pi_1(S^2/\theta) = \{\ [i], [\alpha],\ |\ [\alpha] \star [\alpha] = [i]\ \} \tag{2.17}$$

Under the homomorphism:

$$[i] \to 0, \qquad [\alpha] \to 1 \tag{2.18}$$

this becomes the additive group Z_2 of integers modulo 2. This is an example of a *finite* homotopy group: $\pi_1(S^2/\theta) = Z_2$.

(v) $T^2 = S^1 \times S^1$. This is the 2-torus. As visualized in our own 3-dimensional world, this looks like Fig. 2.8.

To find its fundamental group, we use the theorem stated earlier, about π_1 of a product space. Thus we have $\pi_1(T^2) = Z \oplus Z$. Similarly, for the n-torus $T^n = S^1 \otimes S^1 \otimes \ldots \otimes S^1$ (n times), we have $\pi_1(T^n) = Z \oplus Z \oplus \ldots \oplus Z$ (n times).

For the two-torus, a homotopy class is labelled by a pair of integers (n_1, n_2) which described the number of times a loop winds around the first and second S^1 respectively.

Exercise: Describe some loops on the torus which lie in the homotopy classes (1,0), (0,1), (1,1).

Exercise: Find a space S with $\pi_1(S) = Z \oplus Z_2$ (recall the theorem on π_1 of a

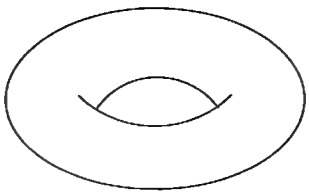


Figure 2.8: The two-torus $S^1 \times S^1$.

product space).

(vi) $\mathbb{R}^2 - B^2$: this is the plane with a disc cut out of it. We can find its π_1 directly, but instead let us show that this space is of the same homotopy type as S^1. This is a nice example of two spaces which are *not* homeomorphic but *are* of the same homotopy type.

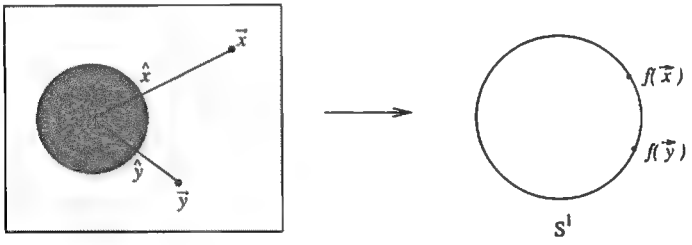


Figure 2.9: $\mathbb{R}^2$ minus a disc is of the same homotopy type as S^1.

This is illustrated in Fig. 2.9. Define $f : \mathbb{R}^2 - B^2 \to S^1$ to be the map which sends every point $\vec{x} \in \mathbb{R}^2 - B^2$ to the corresponding unit vector $\hat{x}$, which lies on S^1. In other words, simply project the point down along the line joining it to the origin until it reaches the unit circle, which defines its image in S^1. The reverse map $g : S^1 \to \mathbb{R}^2 - B^2$ takes all points of the circle S^1 and puts them on the boundary of the disc removed from $\mathbb{R}^2$. Clearly this is a homotopy of topological spaces. We can image "continuously shrinking" $\mathbb{R}^2 - B^2$ down to the circle. Thus $\pi_1(\mathbb{R}^2 - B^2) = \pi_1(S^1) = Z$.

(vii) $\mathbb{R}^2$ minus two discs (see Fig.2.10).

We will show that π_1 of this space is *non-abelian*. Consider the two loops α, β based at x_0 as in the first figure of Fig. 2.10. Clearly they are not homotopic to each other, so $[\alpha] \neq [\beta]$. Now consider the products $[\alpha] \star [\beta] = [\alpha \star \beta] = \gamma$ and $[\beta] \star [\alpha] = [\beta \star \alpha] = \delta$ in the same figure. It is evident that $\gamma \neq \delta$, and therefore $[\alpha] \star [\beta] \neq [\beta] \star [\alpha]$. Thus $\pi_1(S)$ is a non-Abelian group in this case.

This group is known as the free group on two generators, which simply means its elements are abstract products of two independent generators arranged in any order, with different orderings representing independent elements

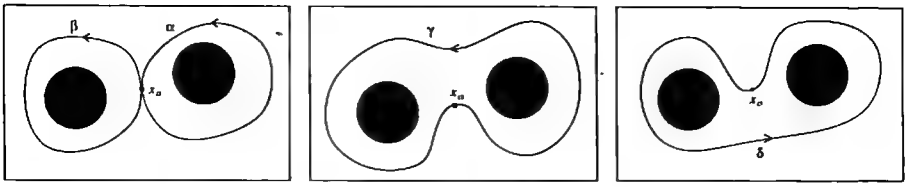


Figure 2.10: Two based loops β, α; a loop γ in the class $[\alpha \star \beta]$; a loop δ in the class $[\beta \star \alpha]$ (in $\mathbb{R}^2$ with two discs removed).

of the group.

To see that π_1 can be non-Abelian in general, consider multiplication of classes of based loops $[\alpha], [\beta]$ in the two possible orders as depicted in Fig. 2.11. Keeping the base point fixed, one cannot in general deform the second diagram

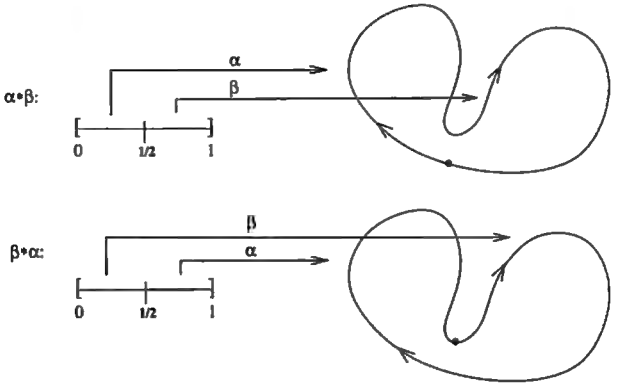


Figure 2.11: The product of loops α, β in different orders. Clearly the two products are not deformable into each other.

of Fig. 2.11 into the first. This is the basic reason why $\pi_1(S)$ is non-Abelian in general.

2.4 Higher Homotopy Groups

We have seen that the fundamental group $\pi_1(S)$ is the collection of homotopy classes of continuous maps from the closed interval [0, 1] to the space S, with the requirement that the map be "based" at a given point x_0. This requirement is implemented by the condition that the end-points of the closed interval be mapped into a common point in S. In this section, we generalize the notion of loops to higher-dimensional objects mapped into a topological space. This will allow us to define more homotopy groups associated to a given space. Each one

could then potentially capture some new topological information, which would lead us closer to the goal of understanding the topology of a space through the study of homotopy.

A simple generalization of a based loop is obtained by considering maps from the closed unit square $[0,1] \otimes [0,1]$ to S, with the requirement that the entire boundary of the square be sent to a common point in X. Thus consider

$$\alpha : [0,1] \otimes [0,1] \rightarrow S \tag{2.19}$$

given by a continuous function $\alpha(t_1, t_2)$ where t_1 is in the first $[0,1]$ and t_2 is in the second $[0,1]$, such that

$$\alpha(0,t_2) = \alpha(1,t_2) = \alpha(t_1,0) = \alpha(t_1,1) = x_0. \tag{2.20}$$

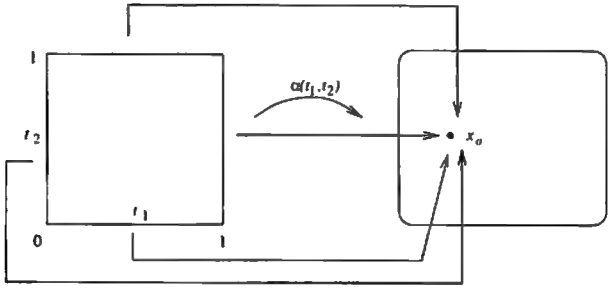


Figure 2.12: A based "two loop" maps the boundary of $[0,1] \times [0,1]$ to a single point x_0.

Such a map may be considered the two-dimensional analogue of the based loops described earlier. Let us call it a 2-*loop*. Now a *homotopy* between two 2-loops $\alpha(t_1,t_2)$ and $\beta(t_1,t_2)$ based at the same point x_0 is defined by:

$$H : [0,1] \otimes [0,1] \otimes [0,1] \rightarrow S \tag{2.21}$$

$H(s,t_1,t_2)$ is a continuous map with the following properties:

$$\begin{aligned} H(0,t_1,t_2) &= \alpha(t_1,t_2) \\ H(1,t_1,t_2) &= \beta(t_1,t_2) \\ H(s,t_1,t_2)\Big|_{(t_1 \text{ or } t_2 = 0 \text{ or } 1)} &= x_0 \text{ for all } s. \end{aligned} \tag{2.22}$$

This map describes a continuous deformation of one 2-loop into another, analogous to what was done for loops in standard homotopy.

All this is easily generalized to "n-loops", though one should keep in mind that "n-loop" is not standard mathematical terminology.

Definition: An *n-loop* in X based at x_0 is a continuous map:

$$\alpha : [0,1]^n \rightarrow X$$

such that

$$\alpha(t_1,\ldots,t_n)\Big|_{(\text{any } t_i=0 \text{ or } 1)} = x_0$$

Here, the nth power of the interval $[0,1]$ denotes its Cartesian product with itself n times.

Definition: A *homotopy of n-loops* $\alpha(t_1,\ldots,t_n)$ and $\beta(t_1,\ldots,t_n)$ is a continuous map

$$H:\ [0,1]^{n+1} \to S$$

such that

$$H(s;t_1,\ldots,t_n)\Big|_{s=0} = \alpha(t_1,\ldots,t_n)$$

$$H(s,t_1,\ldots,t_n)\Big|_{s=1} = \beta(t_1,\ldots,t_n)$$

$$H(s,t_1,\ldots,t_n)\Big|_{\text{any } t_i=0 \text{ or } 1} = x_0 \text{ for all } s$$

This homotopy is an equivalence relation, as one can easily check. With the above definitions, we follow the predictable path of defining the product of generalised n-loops, leading finally to the concept of the nth homotopy group π_n of a space.

Definition: The product of two n-loops α, β is given by

$$\alpha(t_1,\ldots,t_n) \star \beta(t_1,\ldots,t_n) = \alpha(2t_1,t_2,\ldots,t_n), \qquad 0 \le t \le \frac{1}{2}$$

$$= \beta(2t_1-1,t_2,\ldots,t_n), \qquad \frac{1}{2} \le t \le 1$$

In a previous section on loops, we had constructed the inverse of a loop by tracing the loop backwards. Now that we are dealing with "n-loops", an analogous operation can be defined by simply tracing backwards in the first argument t_1 of the loop, leaving the others unchanged.

Definition: The *inverse of an n-loop* α is given by

$$\alpha^{-1}(t_1,\ldots,t_n) = \alpha(1-t_1,t_2,\ldots,t_n)$$

With all these properties it becomes evident that equivalence classes of based n-loops form a group, just as for simple loops.

Definition: The nth *homotopy group* of the space S, with base point x_0, denoted $\pi_n(S,x_0)$, is

$$\pi_n(S,x_0) = \{\ [\alpha] \mid \alpha \text{ is an } n\text{-loop based at } x_0\ \}$$

As before, $[\alpha]$ denotes the equivalence class of all loops homotopic to α, and the group operations are just as for 1-loops. In particular, the identity

element of the group corresponds to the homotopy class of the constant loop, $\alpha(t_1, \ldots, t_n) = x_0$ for all t_i. For path-connected spaces, $\pi_n(S, x_0)$ is independent of x_0 as one would expect.

Example: $\pi_2(S)$ is the group of homotopy classes of 2-loops. A 2-loop is just the image of the 2-sphere S^2 in the given topological space. It is easy to see that in $\mathbb{R}^3$, all 2-spheres can be deformed to a point, so $\pi_2(\mathbb{R}^3) = \{i\}$. On the other hand in $\mathbb{R}^3 - \{0\}$, a 2-sphere enclosing the origin cannot be shrunk. So $\pi_2(\mathbb{R}^3 - \{0\})$ is non-trivial. In fact (although it is not as easy to see intuitively as for the analogous case $\pi_1(\mathbb{R}^2 - \{0\})$), it turns out that

$$\pi_2(\mathbb{R}^3 - \{0\}) = Z \tag{2.23}$$

Before giving more examples, let us state an important theorem.

Theorem: The homotopy groups $\pi_n(S), n \geq 2$, are *Abelian.*

Proof: Instead of giving a formal proof let us use diagrams to indicate how this works. The relevant diagram for π_2 is Fig. 2.13. Let us see how $\alpha \star \beta$

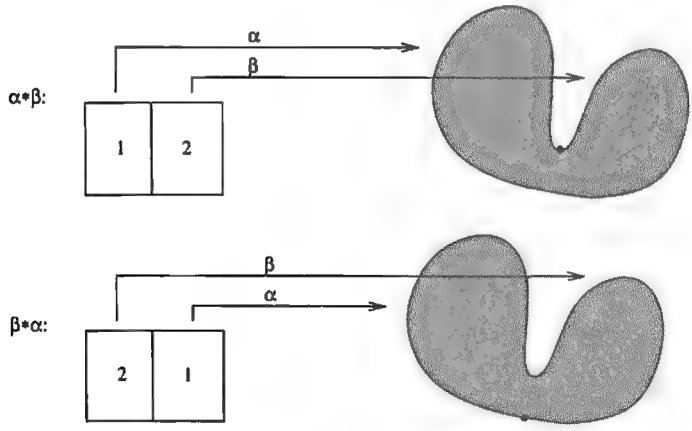


Figure 2.13: The product of two-loops α, β in different orders.

can be homotopically deformed to $\beta \star \alpha$ in this case. Note that α and β are constrained to map all the boundary points to x_0. Thus the product $\alpha \star \beta$ maps the boundary of the rectangle to x_0. The trick now is to continuously modify the *parametrisation* of the map $\alpha \star \beta$ while leaving its *image* in the topological space unchanged. For a 1-loop we could only speed up or slow down the "traversal speed" in terms of the parameter t, of each part of the loop. But here, with two parameters, we can do something more drastic.

We start by homotopically deforming $\alpha \star \beta$ to map many more points in the rectangle to x_0, namely, all those within some distance of the boundary. This is illustrated in Fig. 2.14, where the whole shaded region is mapped to x_0. This

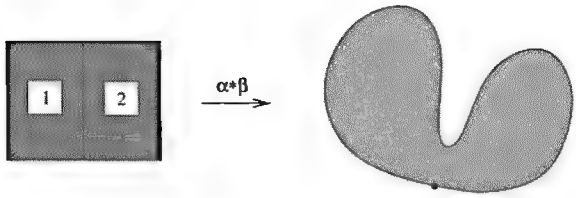


Figure 2.14: The map $\alpha^*\beta$ deformed so that the entire shaded region is mapped to the same point x_0.

operation is continuous as long as the regions marked 1 and 2, over which the map is non-constant, are non-empty regions.

Next, perform the sequence of operations shown in Fig. 2.15. Clearly all the

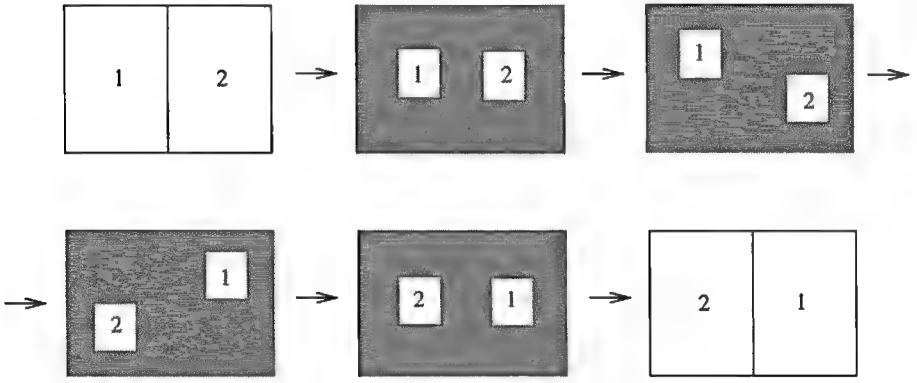


Figure 2.15: Proof that π_2 is Abelian.

operations are continuous and during the whole process, the image of the map remains the same. But at the end we have managed to change the parametrisation of the map from that of $\alpha * \beta$ to that of $\beta * \alpha$. Thus the two 2-loops are homotopic and belong in the same equivalence class. Hence,

$$[\alpha] * [\beta] \equiv [\alpha * \beta] = [\beta * \alpha] \equiv [\beta] * [\alpha] \tag{2.24}$$

and we have shown, as desired, that $\pi_2(S)$ is an Abelian group.

Exercise: Convince yourself that the same procedure does not work for π_1, and that it works for all $n \geq 2$.

To conclude this section, we state without proof a few useful results which tend to occur in problems of physical interest.

(i) $\pi_n(S^n) = Z$

(ii) $\pi_m(S^n) = 0, m < n$

(iii) $\pi_3(S^2) = Z$

The interested reader should consult the mathematical literature for proofs of these and many more results on homotopy groups.

All homotopy groups $\pi_n(S)$ of a topological space are invariant under homeomorphisms, and hence should be thought of as topological invariants characterising the space. They cannot be changed except by discontinuous (non-homeomorphic) deformations of the space.

Chapter 3

Differentiable Manifolds I

Starting with a topological space, it is sometimes possible to put additional structure on it which makes it *locally* look like Euclidean space $\mathbb{R}^n$ of some dimension n. On this space, by imposing suitable requirements, it will then be possible to differentiate functions. The space so obtained is called a *differentiable manifold*. This notion is of central importance in General Relativity, where it provides a mathematical description of spacetime.

3.1 The Definition of a Manifold

Among the topological spaces we have studied, the Euclidean space $\mathbb{R}^n = \mathbb{R} \otimes \mathbb{R} \otimes \ldots \otimes \mathbb{R}$ is special. Besides being a metric space and hence a topological space with the metric topology, it has an intuitive notion of "dimension", which is simply the number n of copies of $\mathbb{R}$.

Let us concentrate on two simple cases: $\mathbb{R}$, the (one-dimensional) real line, and $\mathbb{R}^2$, the (two-dimensional) plane. If the dimension of a space is to be topologically meaningful, two different $\mathbb{R}^n$, $\mathbb{R}^m$ should not be homeomorphic to each other, for $m \neq n$. This is indeed the case.

Let us demonstrate that $\mathbb{R}$ is not homeomorphic to $\mathbb{R}^2$. Assume the contrary: suppose there is a homeomorphism (recall that this means a 1-1 continuous function with continuous inverse) $f : \mathbb{R} \to \mathbb{R}^2$. Consider the restriction of this function: $f : \mathbb{R} - \{0\} \to \mathbb{R}^2 - \{0\}$ where we have deleted the point $\{0\}$ from $\mathbb{R}$, and its image under f, which we define to be the origin, from $\mathbb{R}^2$.

Now if $f : \mathbb{R} \to \mathbb{R}^2$ is a homeomorphism then so is $f : \mathbb{R} - \{0\} \to \mathbb{R}^2 - \{0\}$. But $\mathbb{R} - \{0\}$ is disconnected, while $\mathbb{R}^2 - \{0\}$ is clearly still connected. Since connectivity is invariant under homeomorphism, this is a contradiction. Thus the original homeomorphism $f : \mathbb{R} \to \mathbb{R}^2$ does not exist, and $\mathbb{R}$ is not homeomorphic to $\mathbb{R}^2$. Similarly, one finds that $\mathbb{R}^n$ and $\mathbb{R}^m$ are not homeomorphic for $m \neq n$. So "dimension" is indeed a topological property.

Now consider two more examples of topological spaces: the circle,

$$S^1 = \{ (x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \} \tag{3.1}$$

and the 2-sphere,

$$S^2 = \{ (x,y,z) \in \mathbb{R}^2 \mid x^2 + y^2 + z^2 = 1 \} \tag{3.2}$$

S^1 and S^2 inherit the relative topology as subsets of $\mathbb{R}^2$ and $\mathbb{R}^3$ with the usual topology. Intuitively, we see that S^1 and S^2 are 1-dimensional and 2-dimensional respectively. But in what sense can we make this precise? Certainly they are not homeomorphic to $\mathbb{R}$ and $\mathbb{R}^2$ respectively.

Theorem: S^1 is not homeomorphic to $\mathbb{R}$.

Proof: Assume a homeomorphism $f : S^1 \to \mathbb{R}$. Delete a point and obtain $f : S^1 - \{0\} \to \mathbb{R} - \{0\}$. But $S^1 - \{0\}$ is connected while $\mathbb{R} - \{0\}$ is not. So S^1 and $\mathbb{R}$ are not homeomorphic. Similarly $S^2, \mathbb{R}^2$ are not homeomorphic. (In this case, both remain connected after deleting a point, but S^2 also remains simply connected, while $\mathbb{R}^2$ does not.)

Then what is the common property of S^1 and $\mathbb{R}$, or S^2 and $\mathbb{R}^2$, which suggests that they have the same dimension? It is that they are *locally* homeomorphic. Let us now make this precise.

Definition: Let M be a topological space. A *chart* C on M is a collection (U, ϕ, n) such that:
(i) U is an open set of M.
(ii) ϕ is a homeomorphism: $U \subset M \to \phi(U) \subset \mathbb{R}^n$. (Thus, $\phi(U)$ is open in $\mathbb{R}^n$.)
(iii) n is a positive integer, called the *dimension* of the chart C.

In plain language, a chart is a piece of the topological space, with a homeomorphism equating it to a piece of some Euclidean space.

It is evident that $\mathbb{R}^n$ comes with a preferred choice of coordinates, called *Cartesian coordinates*, namely the n real numbers specifying each point. Now given a chart, each point on it can be labelled by considering the image of this point in $\mathbb{R}^n$ and then using the Cartesian coordinates of that point.

Definition: Given a chart $C = (U, \phi, n)$, the *coordinates* of a point $p \in U \subset M$ are the Cartesian coordinates in $\mathbb{R}^n$ of the point $\phi(p) \in \phi(U) \subset \mathbb{R}^n$. When we allow the points p to vary over U, the Cartesian components of $\phi(p)$ in $\mathbb{R}^n$ define a set of n real-valued functions on $U \in M$ called *coordinate functions*.

The concept of chart has enabled us to locally map a piece of a topological space to $\mathbb{R}^n$. A function on this piece of the topological space can then be *differentiated* by simply differentiating, in the familiar sense, the components of the coordinate functions defined above.

Now we would like to cover the entire topological space with charts so that we will be able to differentiate functions over the entire topological space.

Necessarily the open sets U will overlap with each other (since a topological space, unless it is disconnected, cannot be expressed as a union of disjoint open sets). The notion of differentiability which the charts embody should then be the same for two overlapping charts in their region of overlap. This leads us to the following definition.

Definition: Let $C_1 = (U_1, \phi_1, n), C_2 = (U_2, \phi_2, n)$ be two charts on M. Then C_1 and C_2 are "C^∞-compatible" or simply "compatible" if:
(i) $U_1 \cap U_2 = \phi$, or
(ii) the maps $\phi_1 \circ \phi_2^{-1} : \phi_2(U_1 \cap U_2) \to \phi_1(U_1 \cap U_2)$ and $\phi_2 \circ \phi_1^{-1} : \phi_1(U_1 \cap U_2) \to \phi_2(U_1 \cap U_2)$ are infinitely differentiable (C^∞) functions.

Note that $\phi^1 \circ \phi_2^{-1}$ and $\phi_2 \circ \phi_1^{-1}$ are continuous maps from subsets of $\mathbb{R}^n$ to subsets of $\mathbb{R}^n$. This can be represented as in Fig. 3.1.

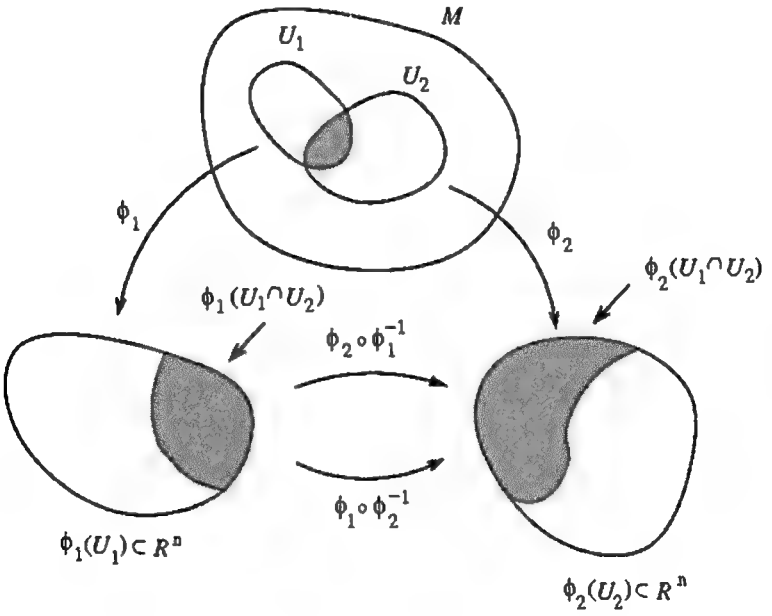


Figure 3.1: Definition of a manifold.

Thus, two charts are compatible if the functions between subsets of $\mathbb{R}^n$ shown in the diagram can be differentiated an arbitrary number of times. Now we are in a position to cover the space with charts.

Definition: A C^∞-*atlas* $\mathcal{U}$ on a topological space M is a family of charts $(U_\alpha, \phi_\alpha, n)$ which covers M, so that $\cup_\alpha U_\alpha = M$, such that all the charts in the family are mutually C^∞-compatible.
Definition: Two C^∞-atlases are said to be *compatible* (with each other) if every

chart of one atlas is compatible with every chart of the other atlas.

Theorem: For atlases, C^∞-compatibility is an equivalence relation.

The proof is useful in giving the reader an idea what an atlas really means, and is left as an exercise since it is straightforward. On the way, the alert reader will discover that compatibility of individual charts is *not* an equivalence relation.

Basically, an atlas is a collection of charts that covers the space, such that we can go from one to the other (this really means going from one subset of $\mathbb{R}^n$ to another) by a differentiable function. The charts and their collections into atlases form the building blocks of the concept of "differentiable structure" on a topological space. A space with such a structure will be called a differentiable manifold.

Definition: A *differentiable structure* of class C^∞ on a topological space M is an equivalence class of C^∞-compatible atlases on M.

It is always easier to study equivalence classes if we can find a unique representative of each class in some way, since this can then be used to label the class. We may pick a unique atlas out of each equivalence class of compatible atlases as follows: take the union of all atlases in a class and call it the *maximal atlas*. Then a differentiable structure is just a choice of maximal atlas $\mathcal{U}$ on M. This choice therefore labels the differentiable structure.

We are finally in a position to define a manifold.

Definition: $A(C^\infty)$ *differentiable manifold* is the pair $(M,\mathcal{U})$ of a Hausdorff topological space M and a C^∞ differentiable structure $\mathcal{U}$.

From this it follows that a differentiable manifold has the following properties:

(i) It is locally Euclidean.

(ii) It is locally compact (every point $x \in M$ has a compact neighbourhood).

(iii) An open subset $U \in M$ is itself a differentiable manifold, called an "open submanifold".

(iv) The product of two manifolds is well-defined (using Cartesian products).

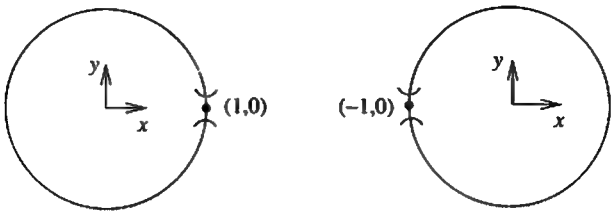
In general terms, the definition of differentiable manifold says that each local region of the space looks like a local region of $\mathbb{R}^n$, and that the many local regions on M are "patched up" by piecing together local regions on $\mathbb{R}^n$ using differentiable functions.

Examples:

(i) $\mathbb{R}^n$ is obviously a differentiable manifold. All of it can be covered by a single coordinate chart, and the coordinate map is the identity.

(ii) S^1 is a manifold. We cannot choose all of S^1 to be a single chart, since this is not homeomorphic to any open set of any $\mathbb{R}^n$. So let us cover it with two overlapping charts. Since S^1 is defined as a subset $x^2+y^2=1$ of $\mathbb{R}^2$, it inherits

the relative topology. A basis is given by the usual open intervals.



$U_1 : \{ (x,y) \in S^1 \mid (x,y) \neq (1,0) \}$ $\quad$ $U_2 : \{ (x,y) \in S^1 \mid (x,y) \neq (-1,0) \}$

Figure 3.2: Coordinate charts for the circle, S^1.

Now to define S^1 as a manifold take two open subsets of S^1, as shown in Fig. 3.2. One consists of the entire space minus the point $(1,0)$ while the other consists of the entire space minus the point $(-1,0)$. Next define the homeomorphisms:

$$\begin{aligned} \phi_1 &: U_1 \to \mathbb{R}, \qquad \phi_1(x,y) = \tan^{-1}\frac{y}{x} = \theta_1, \qquad \theta_1 \in (0, 2\pi) \\ \phi_2 &: U_2 \to \mathbb{R}, \qquad \phi_2(x,y) = \tan^{-1}\frac{y}{x} = \theta_2, \qquad \theta_2 \in (-\pi, \pi) \end{aligned} \tag{3.3}$$

Now we must look at the overlaps $U_1 \cap U_2$. There are two such regions: the upper half of the circle and the lower half of the circle. On the upper region we have:

$$\begin{aligned} \phi_1(U_1 \cap U_2)_{\text{upper}} &= (0, \pi) \\ \phi_2(U_1 \cap U_2)_{\text{upper}} &= (0, \pi) \end{aligned} \tag{3.4}$$

and $\phi_2 \circ \phi_1^{-1} : (0,\pi) \to (0,\pi)$ is the identity map:

$$\phi_2 \circ \phi_1^{-1}(\theta_1) = \theta_1. \tag{3.5}$$

on the lower region, we find:

$$\begin{aligned} \phi_1(U_1 \cap U_2)_{\text{lower}} &= (\pi, 2\pi) \\ \phi_2(U_1 \cap U_2)_{\text{lower}} &= (-\pi, 0) \end{aligned} \tag{3.6}$$

So, $\phi_2 \circ \phi_1^{-1} : (\pi, 2\pi) \to (-\pi, 0)$ is the map:

$$\phi_2 \circ \phi_1^{-1}(\theta_1) = \theta_1 - 2\pi \tag{3.7}$$

Both θ_1 and $\theta_1 - 2\pi$ are functions that can be differentiated arbitrarily many times (in this simple case they are just linear functions!). This shows that S^1 is a differentiable manifold.

(c) S^n, the n-sphere, defined by $\sum_{i=1}^{n+1} x_i^2 = 1$ in $\mathbb{R}^{n+1}$, is a differentiable manifold. This can be shown using stereographic projection. Let us work this out pictorially and then algebraically.

On S^2, choose the open sets (in the topology induced from $\mathbb{R}^3$) to be $U_1 = S^2 - \{\text{north pole}\}$, $U_2 = S^2 - \{\text{south pole}\}$. We map these onto $\mathbb{R}^2$ by stereographic projection as illustrated in Fig. 3.3. Each point $p \in S^2 - \{\text{north pole}\}$ goes to a corresponding point $p' \in \mathbb{R}^2$. For the other chart U_2, projection is done through the south pole. The projection is clearly a homeomorphism. In this case the image of the projection in $\mathbb{R}^2$ is the whole of $\mathbb{R}^2$ rather than a subset, which is fine since the whole space is always an open set of its topology. Thepicture enables one to visualise the maps ϕ_1, ϕ_2 and convince oneself that $\phi_1 \circ \phi_2^{-1}, \phi_2 \circ \phi_1^{-1}$ are C^∞.

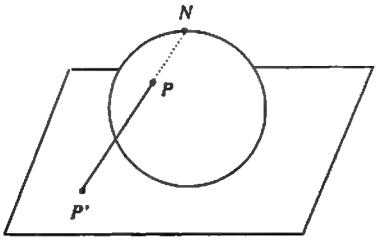


Figure 3.3: Stereographic projection of S^2 onto $\mathbb{R}^2$.

This pictorial construction is hard to generalise to S^n for $n > 2$ but only because our imagination does not work in high dimensions! However, the algebraic description of stereographic projections does work in all dimensions. Let us work it out explicitly for S^1, since the construction is analogous for all S^n. Indeed, this will give us another way of specifying coordinate charts on S^1, besides the one described in a previous example.

From Fig. 3.4, the point $p \in S^1$ is mapped onto the point:

$$\phi_1(p) = \frac{2a}{\tan\frac{\theta}{2}} \in \mathbb{R}^1. \tag{3.8}$$

This is defined for $\theta \neq 0$, that is, $p \in M_1$ where M_1 is the chart excluding $\theta = 0$. For the opposite stereographic projection (upwards from the south pole), we will get

$$\phi_2(p) = \frac{2a}{\cot\frac{\theta}{2}} = 2a\tan\frac{\theta}{2} \tag{3.9}$$

which is defined for $\theta \neq \pi$, hence for $p \in M_2$.

Now consider $\phi_2 \circ \phi_1^{-1} : \mathbb{R} \to \mathbb{R}$. We have:

$$\begin{aligned} \phi_1^{-1}(0, 2\pi) &= 2\cot^{-1}\tfrac{x}{2a} \\ \phi_2 \circ \phi_1^{-1} &= 2a\tan\left(\cot^{-1}\tfrac{x}{2a}\right) \\ &= \tfrac{4a^2}{x} \end{aligned} \tag{3.10}$$

This is the inversion map: $\mathbb{R} - \{0\} \to \mathbb{R} - \{0\}$ which is infinitely differentiable. The map $\phi_1 \circ \phi_2^{-1}$ works similarly.

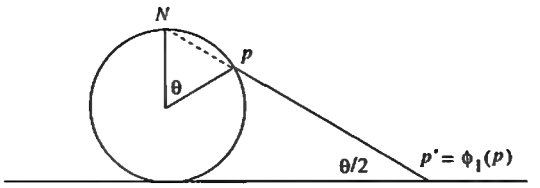


Figure 3.4: A simpler case: stereographic projection of S^1 onto $\mathbb{R}$.

3.2 Differentiation of Functions

One useful outcome of our definition of differentiable manifolds is that we can now differentiate functions from one manifold X to another manifold Y. The idea is simple. Take a chart on X and map it to an open set of some $\mathbb{R}^n$. Take the image of this chart in Y under the given function and map it to an open set of some $\mathbb{R}^m$. Then composing the coordinate maps and the given function appropriately, we get a function from a region of $\mathbb{R}^n$ to a region of $\mathbb{R}^m$, which we can then differentiate.

Let us work this out. Take two manifolds X and Y and a function $f : X \to Y$. The coordinate map ϕ takes an open set U of X to an open set $\phi(U) \subset \mathbb{R}^n$. Similarly the coordinate map ψ takes an open set V of Y and maps it to an open set $\psi(V) \subset \mathbb{R}^m$. The situations is illustrated in Fig. 3.5. Clearly the function $\psi \circ f \circ \phi^{-1}$ takes us from $\phi(U)$ to $\psi(V)$. If this function is C^∞ (infinitely differentiable) for all U, V, then we say that $f : X \to Y$ is a C^∞ map.

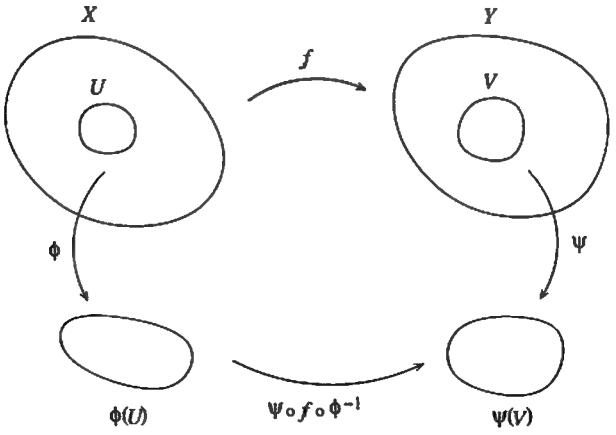


Figure 3.5: Differentiation of a function $f : X \to Y$.

In the previous section we defined two kinds of coordinates on the topolog-

ical space S^1, one by a simple map of segments and the other by stereographic projection. Have we defined two *different* differentiable manifolds? To answer this, we need to define what we mean by equivalence of two differentiable manifolds. Clearly they must, for a start, be homeomorphic as topological spaces. But the differentiable nature must also be preserved.

Definition: Two manifolds X, Y are said to be *diffeomorphic* if there exists a homeomorphism $f : X \to Y$ such that f is a C^∞ function with a C^∞ inverse. f is called a *diffeomorphism.*

If two differentiable manifolds are diffeomorphic, we think of them as the same manifold.

As an example, let us reconsider S^1, which was discussed in Examples (ii) and (iii) of the previous section. In (ii), the coordinate on one patch was labelled θ. In (iii), the coordinate was $\frac{2a}{\tan\frac{\theta}{2}}$. Thus we have a map,

$$\theta \to \frac{2a}{\tan\frac{\theta}{2}} \tag{3.11}$$

which (on $(0, \pi)$) is C^∞ with a C^∞ inverse, hence it is a diffeomorphism and the two descriptions are diffeomorphic to each other. Thus the two ways in which we defined S^1 gave rise to the *same* differentiable manifold.

Theorem: $S^n, n \leq 6$, admits a unique differentiable structure. But for $n \geq 7$ there are many possible inequivalent differentiable structures. (This is a highly non-trivial result!)

3.3 Orientability

For 2-dimensional manifolds visualised in 3 dimensions, we have the intuitive notion of two different *sides* of the manifold. For $\mathbb{R}^2$, we often say "out of plane" and "into the plane", while for S^2 there is an "outside" and an "inside". This is an embryonic form of the notion of orientation and orientability, which we will define precisely in a moment.

First, consider the strange example of a two-dimensional manifold which has only *one* side. This is called the Möbius strip. It is constructed by taking the rectangle $\{ (x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq a,\ 0 \leq y \leq b \}$ and identifying a pair of opposite sides with a "twist", in other words, joining them so that the arrows match in Fig. 3.6. If we now take a normal vector and call it "outward", then on transporting it once around the strip, it comes back pointing "inward", namely its direction has been reversed. There is only one "side" to the Möbius strip!

To formalize this notion for general manifolds, consider two overlapping coordinate patches (U_α, ϕ_α) and (U_β, ϕ_β). Consider the map,

$$\phi_\alpha \circ \phi_\beta^{-1} : \quad \phi_\beta(U_\alpha \cap U_\beta) \to \phi_\alpha(U_\alpha \cap U_\beta) \tag{3.12}$$

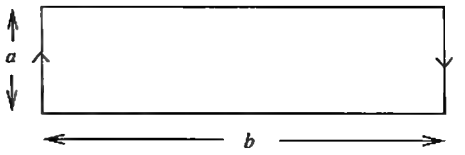


Figure 3.6: How to make a Möbius strip.

As a map $\mathbb{R}^n \to \mathbb{R}^n$, we can describe this by a function

$$f : x^i \to y^i(x_1, \ldots, x_n) \tag{3.13}$$

The Jacobian of this map is the determinant:

$$J = \left\| \frac{\partial y_i}{\partial x_j}(x) \right\| \tag{3.14}$$

Definition: A manifold is *orientable* if we can choose an atlas on it such that on every overlap $U_\alpha \cap U_\beta$ of charts, the map $f = \phi_\alpha \circ \phi_\beta^{-1}$ has a positive Jacobian determinant.

Example: The cylinder is orientable. Define the cylinder as $S^1 \times [0,1]$. Choose the coordinate patches to be $U_1 \times (0,1)$ and $U_2 \times (0,1)$, where U_1, U_2 were the coordinate patches on S^1 in Example (b) above. Define the homeomorphisms $\phi_1 : U_1 \cap U_2 \to \mathbb{R}$, $\phi_2 : U_1 \cap U_2 \to \mathbb{R}$ (see Fig. 3.7).

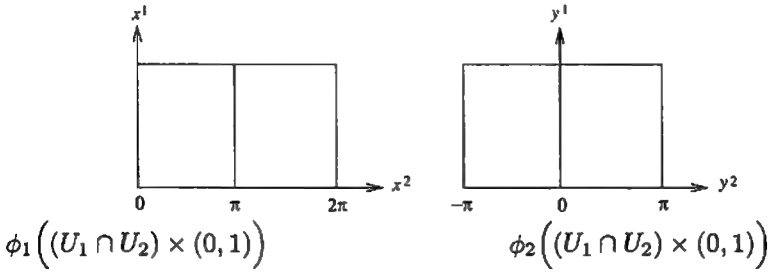


Figure 3.7: Coordinate charts for a cylinder.

Now the function $\phi_2 \circ \phi_1^{-1}$ is:

$$\begin{aligned} \phi_2 \circ \phi_1^{-1} : \ & (x^2, x^2) \to (x^1, x^2) && 0 < x^1 < \pi \\ & (x^1, x^2) \to (x^1 - 2\pi, x^2) && \pi < x^1 < 2\pi \end{aligned} \tag{3.15}$$

In other words $y^i = y^i(x)$ is: $(y^1, y^2) = (x^1, x^2)$ or $(y^1, y^2) = (x^1 - 2\pi, x^2)$. In each region, we have $\frac{\partial y^i}{\partial x^j} = \delta^i_j$ and $J = 1$, so the cylinder is orientable.

For the Möbius strip, we define the manifold by specifying different functions on the overlaps:

$$\begin{aligned} \phi_2 \circ \phi_1^{-1} : \ & (x^1, x^2) \to (x^1, x^2) && 0 < x^1 < \pi \\ & (x^1, x^2) \to (x^1 - 2\pi, 1 - x^2) && \pi < x^1 < 2\pi \end{aligned} \tag{3.16}$$

Then $J = 1$ and $J = -1$ in the two segments respectively. This proves that the Möbius strip is not orientable.

3.4 Calculus on Manifolds: Vector and Tensor Fields

One of the main purposes of defining a differentiable manifold is that, given its local equivalence to $\mathbb{R}^n$, we can try to do calculus on the manifold just as we do in $\mathbb{R}^n$. All we have to do is use coordinate charts to get from the manifold to a subset of $\mathbb{R}^n$, after which we do the usual calculus there. The only subtlety is that, since the manifold in general requires more than one chart to cover it, we have to learn how to transfer our calculations across charts.

Let us work with real-valued functions on a manifold M, namely maps $f : M \to \mathbb{R}$. If not otherwise specified, "function on M" will always means a real-valued function. A simple example is the temperature at each point on the surface of the earth, which defines a real-valued function $f : S^2 \to \mathbb{R}$.

Next given such a function f, consider, in a coordinate patch, the associated map $\mathbb{R}^n \to \mathbb{R}$ (see Fig. 3.8). We will frequently use the notation x^i for coordinates on $\phi_\alpha(U_\alpha)$. Also, the function $f \circ \phi_\alpha^{-1}$ in the given coordinate system is denoted as:

$$f : \quad x^i \in \mathbb{R}^n \to f(x^i) \in \mathbb{R}. \tag{3.17}$$

In a different coordinate system, we will have a different function

$$f' : \quad y^i \in \mathbb{R}^n \to f'(y^i) \in \mathbb{R} \tag{3.18}$$

But these must produce the same map $f : M \to \mathbb{R}$ for any point in M which lies in the overlap of U_α, U_β. Thus on the overlap, the relation is

$$f'(y^i) = f(x^i) \tag{3.19}$$

with $y^i = y^i(x)$ given by $\phi_\beta \circ \phi_\alpha^{-1}$.

Now to differentiate $f : M \to \mathbb{R}$, we simply carry this out on $f(x) : \mathbb{R}^n \to \mathbb{R}$. If $f(x)$ is C^∞, then $f : M \to \mathbb{R}$ is said to be C^∞. Differentiating $f(x) : \phi_\alpha(U_\alpha) \subset \mathbb{R}^n \to \mathbb{R}$, we get a set of functions:

$$\frac{\partial f}{\partial x^i} : \quad \phi_\alpha(U_\alpha) \subset \mathbb{R}^n \to \mathbb{R} \qquad i = 1, \ldots, n. \tag{3.20}$$

Though the function f was coordinate invariant on overlaps, as it must be, we now see that its derivative does not enjoy the same property. Indeed,

$$f'(y) = f(x) \implies \frac{\partial f'}{\partial y^i} = \frac{\partial f(x)}{\partial y^i} = \frac{\partial x^j}{\partial y^i}\frac{\partial f(x)}{\partial x^j} \neq \frac{\partial f(x)}{\partial x^i} \tag{3.21}$$

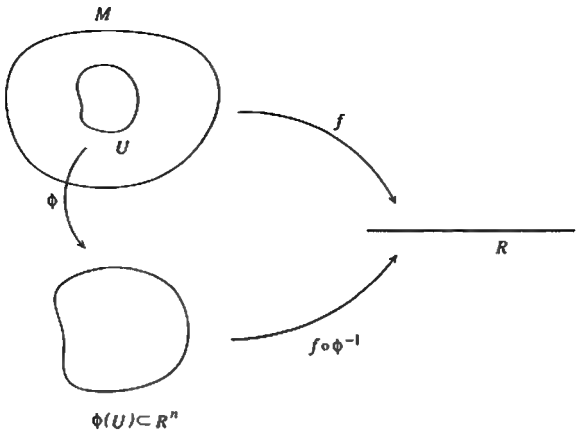


Figure 3.8: Differentiation of a real-valued function.

(Here and below, summation is implied on repeated indices.) Instead of coordinate invariance, we realize that derivatives of functions possess a more complicated transformation property across patches. This may be thought of as "covariance", in other words, a specific transformation law relating the same object in different coordinate patches.

To understand this better, consider a parametrised curve $p(t), p \in M$. This is just some one-dimensional submanifold of M, whose points are labelled by a continuous parameter t. In some given patch, this goes to a curve $x^i(t) \in \mathbb{R}^n$. The tangent vector at t_0 to a curve $x^i(t) \subset \mathbb{R}^n$ is well-known, from elementary geometry, to be $dx^i/dt\,|_{t=t_0}$. (One can think of the tangent vector as measuring the "rate of motion" of a point along the curve, as a function of the "time" t, while it also specifies the instantaneous direction of motion.) But here, dx^i/dt in $\phi_\alpha(U_\alpha)$ describes some property of the original curve $p(t) \subset M$. Accordingly, we *define* this to be the tangent vector to the curve on M, *in the given coordinate patch.*

Definition: The *tangent vector* to a curve $p(t) \subset M$ in a coordinate patch M is dx^i/dt where $x^i(t)$ are the coordinates of the image $\phi_\alpha(p(t))$.

On the overlap of two patches, we have

$$\begin{aligned} x^i &\to y^i(x) \\ \tfrac{dx^i}{dt} &\to \tfrac{dy^i}{dt} = \tfrac{\partial y^i}{\partial x^j}\,\tfrac{dx^j}{dt} \end{aligned} \tag{3.22}$$

Thus given a tangent vector in one patch, the above rule tells us how to relate it to the tangent vector in another overlapping patch. This gives an invariant meaning to the tangent vector. We may say that the tangent vector is the *collection* of objects dx^i/dt in each patch x^i, related on overlaps by the above rule. This is really what we meant by "covariance" above.

Tangent vectors at a point form a *vector space*. If we have two curves $p(t), q(t) \subset M$ passing through the same point at t_0, then we can take the sum of the tangent vectors to each curve at t_0 to get a new vector in $\mathbb{R}^n$, denoted

$$a^i(x_0) = \left[\frac{dx^i}{dt}(p(t)) + \frac{dx^i}{dt}(q(t))\right]_{t=t_0} \tag{3.23}$$

This defines a straight line in $\mathbb{R}^n$:

$$x^i(t) = x_0^i + a^i(x_0)t \tag{3.24}$$

where x_0 is the image of the point $p(0)$ on the manifold.

This straight line can be mapped back to M, to give a new curve on M around the point where $p(t)$, $q(t)$ intersect, using ϕ_α^{-1}. Thus, the sum of tangent vectors to two curves passing through a point can be regarded as a tangent vector to a third curve passing through the same point.

Definition: The vector space of tangent vectors at a point $p \in M$ is called the *tangent space* $T_p(M)$ at p.

Note that any collection of numbers a^i and a coordinate system x^i define a tangent vector. Here is the construction:

(i) Pick a point $x_0^i \in \mathbb{R}^n$ and define any curve $x^i(t)$ in $\mathbb{R}^n$ such that $\left.\frac{dx^i}{dt}\right|_{t=t_0} = a^i$.

(ii) "Lift" the curve back to M using ϕ_α^{-1} for that coordinate system.

(iii) In any other coordinate system $y^i(x)$, define $a'^i = \left.\frac{\partial y^i}{\partial x^j}\right|_{x_0} a^j$ to be the same object in the new coordinates.

Let us now combine the two concepts we have just discussed: differentiation of functions on M, and tangent vectors to curves on M. In this way we will be able to provide a *coordinate invariant* definition of a tangent vector.

Recall how we differentiated a C^∞ function on M along a given curve. If the function on M is f, then along the curve it is $f(p(t))$, and its derivative along the curve is defined to be $\frac{df}{dt}(p(t))$. Clearly this defines a *new* function on the same curve.

Using the coordinate map we can express this as

$$\frac{df}{dt} = \frac{df}{dx^i}\frac{dx^i}{dt}(p(t)) \tag{3.25}$$

Thus, the derivative of f along a curve, at some point, is expressed in terms of its gradient $\partial f/\partial x^i$ in the chosen coordinate system, contracted with the tangent vector at that point. But while both $\partial f/\partial x^i$ and dx^i/dt are coordinate-dependent objects, when contracted together they give df/dt which is manifestly coordinate independent.

Writing

$$\frac{df}{dt} = \left[\frac{dx^i}{dt}(p(t))\frac{\partial}{\partial x^i}\right] f \tag{3.26}$$

we see that $\frac{dx^i}{dt}(p(t))\frac{\partial}{\partial x^i}$ is a coordinate-invariant first order differential operator which maps any C^∞ function on a curve in a manifold into another such function on this curve. At a point $x_0^i = x^i(t_0)$, the operator is $\frac{dx^i}{dt}|_{t=t_0}\frac{\partial}{\partial x^i}$. This carries precisely the same information as the tangent vector $\frac{dx^i}{dt}|_{t_0}$. So we can as well call *this* differential operator a *tangent vector* at T_0.

All we have done is to take our previous definition of tangent vector and contract it with $\partial/\partial x^i$ to make it a *differential operator* on functions. By this process it also becomes manifestly coordinate invariant.

So far we have confined our attention to tangent vectors defined at a fixed point $p \in M$. These can be written $a^i\,\partial/\partial x^i$ for some constants a^i. But now consider the differential operator $X = a^i(x)\,\partial/\partial x^i$, where $a^i(x)$ are C^∞ functions. This operator maps C^∞ functions over a whole coordinate patch into new C^∞ functions on the same patch:

$$X : f \to Xf \equiv a^i(x)\frac{\partial f}{\partial x^i} \tag{3.27}$$

We call X a *vector field*.

Given two functions f and g on M, we have $X(fg) = (Xf)g + f(Xg)$. In fact, this property characterizes vector fields, and we can as well define a vector field this way.

Definition: A *vector field* is a linear map from C^∞ functions to C^∞ functions on a manifold M, satisfying $X(fg) = (Xf)g + f(Xg)$.

Let us now generalise this concept to *tensor fields*. Just as vector fields are linear maps, tensor fields will be defined as multilinear maps from sets of functions to sets of functions. Consider an ordered set of n functions $(f_1, \ldots, f_n)$ on a manifold M. This set forms a vector space under addition in the obvious way. Moreover, we can multiply two elements of this space as follows:

$$(f_1, \ldots, f_n) \star (g_1, \ldots, g_n) = (f_1 g_1, \ldots, f_n g_n)$$

Now a map $X : (f_1, \ldots, f_n) \to X(f_1, \ldots, f_n)$, taking each set of functions in this space to another set, is called *multilinear* if

$$X(a_1 f_1, \ldots, a_n f_n) = a_1 \ldots a_n X(f_1, \ldots, f_n) \tag{3.28}$$

Definition: A *tensor field of rank* n is a multilinear map

$$X : (f_1, \ldots, f_n) \to X(f_1, \ldots, f_n) \tag{3.29}$$

satisfying

$$\begin{aligned} &X((f_1, \ldots, f_n) \star (g_1, \ldots, g_n)) = \\ &(X(f_1, \ldots, f_n)) \star (g_1, \ldots, g_n) + (f_1, \ldots, f_n) \star (X(g_1, \ldots, g_n)) \end{aligned} \tag{3.30}$$

The nth rank tensor fields form a vector space which we denote $T_p(M) \otimes T_p(M) \otimes \ldots \otimes T_p(M)$, or simply $\otimes_n T_p(M)$. This is the n-fold tensor product of the tangent space with itself. Since $\partial/\partial x^i$ form a basis for $T_p(M)$, a basis for the n-fold tensor product space is given by $\partial/\partial x^{i_1} \otimes \partial/\partial x^{i_2} \otimes \ldots \otimes \partial/\partial x^{i_n}$. Thus any $A \in \otimes_n T_p(M)$ can be written

$$A = A^{i_1 i_2 \ldots i_n}(x) \frac{\partial}{\partial x^{i_1}} \otimes \frac{\partial}{\partial x^{i_2}} \otimes \ldots \otimes \frac{\partial}{\partial x^{i_n}} \tag{3.31}$$

Clearly the coefficients $A^{i_1 i_2 \ldots i_n}(x)$ transform like nth rank contravariant tensors, in "physics" terminology, and we call $A \in (T_p(M))^n$ an nth rank tensor field.

Now suppose we have two manifolds M and N and a C^∞ map

$$\phi : M \to N \tag{3.32}$$

This map induces another map, called the *differential map*

$$\phi_* : T_p(M) \to T_{\phi(p)}(N) \tag{3.33}$$

which maps the tangent space to M at p, to the tangent space to N at $\phi(p)$. To define this map, note the following: if $f : N \to \mathbb{R}$ is a C^∞ function on N, then $f \circ \phi : M \to \mathbb{R}$ is a C^∞ function on M (Fig. 3.9).

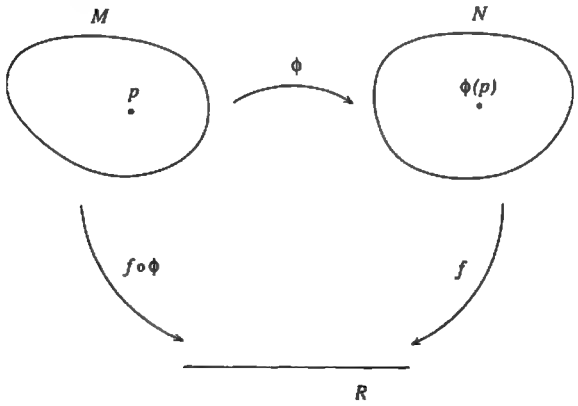


Figure 3.9: How to define the differential map ϕ_*.

A tangent vector acts on functions, as we have discussed above. Given a vector $X \in T_p(M)$, we define $\phi_* X \in T_{\phi(p)}(N)$ to have the same action on $f : N \to \mathbb{R}$ as $X \in T_p(M)$ would have on $f \circ \phi : M \to \mathbb{R}$.

$$(\phi_* X) f = X(f \circ \phi) \tag{3.34}$$

Notice that for given X and f, both sides of this equation are just real numbers. The above equation defines a map $\phi_* : T_p(M) \to T_{\phi(p)}(N)$ by $\phi_* : X \to \phi_* X$.

3.5 Calculus on Manifolds: Differential Forms

Having defined vector and tensor fields on a manifold, we now turn to the definition of their "dual" objects, differential forms. Both tensor fields and differential forms play a crucial role in physics, where we tend to think of their roles as rather similar. However in the context of differentiable manifolds the two have very different meanings, as we will try to bring out in what follows.

Given a vector space V with basis e_i, $i = 1, \cdots, n$, the *dual* vector space V^* is defined to be the vector space generated by the dual basis e^{*i} given by an inner product

$$\langle e^{*i}, e_j \rangle = \delta^i_j \tag{3.35}$$

For finite-dimensional vector spaces, the dual space is isomorphic to the original one. However, the pairing between a vector space and its dual can lead to different properties for the elements of the two spaces. These properties are induced by a space on its dual if we require that the inner product remain invariant under certain transformations. In the present case we want to study the dual of the tangent space, and we will derive the transformation laws across coordinate patches for elements of this space by requiring that the inner product be coordinate invariant.

In general, elements of V and V^* can be expressed in terms of their respective bases:

$$\begin{aligned} &a \in V, && a = \textstyle\sum_j a^j e_j \\ &a' \in V^*, && a' = \textstyle\sum_i a'_i e^{*i} \end{aligned} \tag{3.36}$$

Then, $\langle a', a \rangle = a'_i a^j \langle e^{*i}, e_j \rangle = a'_i a^i$ (summation implied). Thus the pair $a' \in V^*$, $a \in V$ gets mapped onto a real number, or in other words, $\langle \ \ \rangle$ is a bilinear map: $V^* \otimes V \to \mathbb{R}$.

There is an alternative way to look at this. For a given $a' \in V^*$, we have a *linear map*

$$\langle a', \ \rangle : \quad V \to \mathbb{R} \tag{3.37}$$

defined by

$$\langle a', \ \rangle : \quad a \in V \to \langle a', a \rangle \in \mathbb{R} \tag{3.38}$$

Thus the dual vector space to V is the space of *linear functionals* on V.

Let us clarify what we mean by linear functionals. A functional is a map which takes an element of vector space to a number. In our case, clearly $\langle a, \ \rangle$ is a functional, and it has the additional property that:

$$\langle a', \rangle : \lambda a + \mu b \to \lambda \langle a', a \rangle + \mu \langle a', b \rangle \tag{3.39}$$

which is just what we mean by linearity.

An important property of dual spaces is that if we have a map of vector spaces V, W:

$$f : V \to W$$

then this induces a *dual map* in the opposite direction:

$$f^* : W^* \to V^*$$

Let us show this. Given $b' \in W^*$, we want to define $f^*(b') \in V^*$. We can define it as a linear functional on V:

$$\langle f^*(b'), \ \rangle : \quad \langle f^*(b'), a\rangle = \langle b', f(a)\rangle \tag{3.40}$$

The inner product on the left of the equality is clearly between elements of V and V^*, while the one on the right is between elements of W and W^*. Since the map $f : V \to W$ is given, the above equation defines the dual map $f^* : W^* \to V^*$.

In the study of differentiable manifolds, we defined the tangent space $T_p(M)$ at a point p, with basis $\partial/\partial x^i$. Now we will define its dual.

Definition: The space dual to $T_p(M)$ is called the *cotangent space* $T_p^*(M)$.

Denote the dual basis of $T_p^*(M)$ by dx^i. Note that this is a formal symbol for the moment and does *not* represent an infinitesimal amount of anything! Thus there is an inner product,

$$\langle dx^i, \partial/\partial x^j\rangle = \delta^i{}_j \tag{3.41}$$

This is manifestly invariant under local changes of coordinates, if we assign to dx^i the transformation law

$$dx^i \to dy^i = \frac{\partial y^i}{\partial x^j} dx^j \tag{3.42}$$

The notation dx^i serves to remind us that this object *transforms like an infinitesimal distance*. As we have indicated above, this is only notation – but now we understand what motivates it.

Definition: An element of the cotangent space is written

$$\omega = \omega_i \, dx^i$$

and is called a 1-*form* at p.
The coefficient ω_i transforms like a "covariant vector", in "physics" terminology.

As we did in going from vectors to vector fields, we can generalize a form at a single point to a field defined all over the coordinate patch:

$$\omega(x) = \omega_i(x) \, dx^i$$

with $\omega_i(x)$ a set of functions on $\mathbb{R}^n$. Such an object, with the coefficients $\omega_i(x)$ being differentiable functions, is called a *differential 1-form*.

The inner product between basis vectors of the tangent space $T_p(M)$ and the cotangent space $T_p^*(M)$ gives a product between arbitrary elements of these

two spaces. If $X = a^j(x)\,\partial/\partial x^j$ is a vector field and $\omega(x) = \omega_i(x)dx^i$ is a 1-form, then

$$\langle \omega, X\rangle = \omega_i\, a^j\, \langle dx^i, \partial/\partial x^j\rangle = \omega_i\, a^i \tag{3.43}$$

This is coordinate-independent, as it should be. The left hand side is a pairing between a vector field and a form, each of which is coordinate independent, while on the right hand side we find the coefficients (or "components") of these objects, which are coordinate dependent but transform oppositely across patches precisely in such a way that the product is invariant.

We now define the *exterior derivative* on functions on the manifold as a map d defined by:

$$d : f \to df = \frac{\partial f}{\partial x^i} dx^i \tag{3.44}$$

This associates a 1-form to any function f on a manifold. The components of this 1-form are just the derivatives of f along each of the coordinates. Later we will see that the exterior derivative can be defined to act on differential forms and not just functions. We will provide the complete definition at that stage. The exterior derivative plays a fundamental role in the study of differential forms.

If X is an arbitrary vector field, then (as we discussed above) it can be used to map a function to another function by $f \to Xf$. Now we have the equality

$$Xf = \langle df, X\rangle \tag{3.45}$$

which relates the action of X on f with the inner product between X and the exterior derivative of f.

Exercise: Check the above equality.

Just as for the tangent space, we can study tensor products of the cotangent space: $(T_p^*(M))^m$ whose basis is $dx^{i_1} \otimes \ldots dx^{i_m}$. More generally, we can consider elements of the mixed product $(T_p(M))^n \otimes (T_p^*(M))^m$, which look like

$$A = A^{i_1 \ldots i_n}{}_{j_1 \ldots j_m}\, \frac{\partial}{\partial x^{i_1}} \otimes \ldots \otimes \frac{\partial}{\partial x^{i_n}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_m}. \tag{3.46}$$

For a physicist, the components can be thought of as forming a "mixed tensor" which is nth rank contravariant, mth rank covariant.

Now let us return to the exterior derivative d. We can think of functions as "0-forms". Then the operator d acting on a 0-form produces a 1-form. How about the action of d on 1-forms? That is something we have yet to define.

We could perhaps try something like

$$d : \quad \omega = \omega_i\, dx^i \stackrel{?}{\to} \frac{\partial \omega_i}{\partial x^j}\, dx^i \otimes dx^j$$

But this does not work! If ω_i tranforms like the components of a 1-form, $\partial w^i/\partial x^j$

does *not* transform in a covariant way.

Exercise: Check the above statement.

Physicists are familiar with one way to remedy this, which is to introduce an "affine connection" and a "covariant derivative". But that relies on the existence of a Riemannian metric on the manifold, something we have not yet defined. For now our manifolds are equipped only with a differentiable structure. So if we want to define generalizations of 1-forms which transform nicely under coordinate changes, we have to do something different.

Indeed our problem will be solved if we *antisymmetrise* the derivative. We can show that if ω_i transforms like a covariant vector, then $\frac{\partial \omega_i}{\partial x^j} - \frac{\partial \omega_j}{\partial x^i}$ does transform like a covariant 2nd-rank tensor:

$$\begin{aligned}\partial'_i\omega'_j - \partial'_j\omega'_i &= \left(\frac{\partial x^k}{\partial y^i}\frac{\partial}{\partial x^k}\right)\left(\frac{\partial x^l}{\partial y^j}\omega_l\right) - (i \leftrightarrow j) \\ &= \left(\frac{\partial x^k}{\partial y^i}\frac{\partial x^l}{\partial y^j}\partial_k\omega_l + \frac{\partial x^k}{\partial y^i}\frac{\partial y^m}{\partial x^k}\frac{\partial^2 x^l}{\partial y^m \partial y^j}\omega_l\right) - (i \leftrightarrow j) \qquad (3.47)\end{aligned}$$

On antisymmetrising, the second term drops out and we get the desired result. The presence of the second term before antisymmetrising shows explicitly that the result would not transform like a tensor if we did not antisymmetrise.

To keep track of the fact that $\partial_i\omega_j$ must be antisymmetrised, we write

$$d\omega = \partial_i\omega_j\, dx^i \wedge dx^j$$

where $dx^i \wedge dx^j$ is defined to be $\frac{1}{2}(dx^i \otimes dx^j - dx^j \otimes dx^i)$. The object so obtained is called an *antisymmetric 2-form*, or simply a *2-form*.

This motivates us to define the following structure:

Definition: The *wedge product* of two copies of $T^*_p(M)$ is the vector space spanned by the basis $dx^i \wedge dx^j$. In general, the basis $dx^{i_1} \wedge dx^{i_2} \wedge \cdots \wedge dx^{i_n}$ spans the space of *totally* antisymmetric covariant tensor fields. This is denoted $\wedge^n(M)$ and its elements are called *n-forms*. The union of the spaces of all forms, $\wedge(M) \equiv \cup_n \wedge^n(M)$, is called the *exterior algebra* of the manifold M.

We can now complete our definition of the exterior derivative by specifying how it acts on any differential n-form, mapping it to an $n+1$-form.

Definition: The *exterior derivative on n-forms* is a map

$$d : a \in \wedge^n(M) \rightarrow da \in \wedge^{n+1}(M) \qquad (3.48)$$

defined by:

$$\begin{aligned}a &= a_{i_1\cdots i_n}(x)\, dx^{i_1} \wedge \cdots \wedge dx^{i_n} \qquad \rightarrow \\ da &= \left(\frac{\partial}{\partial x^{i_{n+1}}} a_{i_1\cdots i_n}(x)\right) dx^{i_{n+1}} \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_n} \qquad (3.49)\end{aligned}$$

The components of da are totally antisymmetrised, as is manifest in the notation.

Thus we have built up a structure of forms on the manifold, which have rank $0, 1, \cdots d$ where d is the dimension of the manifold M (clearly a totally antisymmetrised object with rank greater than d must vanish). Along with this, we have a map d which takes us from n-forms to $n+1$-forms by antisymmetrised differentiation.

3.6 Properties of Differential Forms

We list a few properties satisfied by differential forms. They are rather easy to prove from the definition and the reader is encouraged to work out all the proofs.
(i) If $a_m \in \wedge^m(M), b_n \in \wedge^n(M)$, then $a_m \wedge b_n = (-1)^{mn} b_n \wedge a_m$. This means that two different forms anticommute if they are both odd, otherwise they commute.
(ii) $d(a_m \wedge b_n) = da_m \wedge b_n + (-1)^m a_m \wedge db_n$. Thus, upto possible signs, the d operator distributes over products of forms just like an ordinary derivative.
(iii) $d^2 = 0$. In other words, acting twice with the d operator on the same object gives 0. We say that d is a *nilpotent* operator. This will turn out to be of fundamental importance. The proof of nilpotence goes as follows:

$$\begin{aligned} d^2 a &= d\left(\frac{\partial}{\partial x^{i_{n+1}}} a_{i_1 \cdots i_n}(x) dx^{i_{n+1}} \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_n}\right) \\ &= \left(\frac{\partial^2}{\partial x^{i_{n+1}} \partial x^{i_{n+2}}} a_{i_1 \cdots i_n}(x)\right) dx^{i_{n+2}} \wedge dx^{i_{n+1}} \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_n} \\ &= 0 \end{aligned} \tag{3.50}$$

since the two derivatives commute, while the wedge product which contracts with them anticommutes.
(iv) The dimension of $\wedge^n(M)$ for a d-dimensional manifold is given by the binomial coefficient $\binom{d}{n} = \frac{d!}{n!(d-n)!}$. This is due to the antisymmetry, and is easy to check. Also, the dimension vanishes for $n > d$ as one would expect because there are not enough indices to antisymmetrise.

While studying the tangent space, we saw that ϕ: $M \to N$ induces a map $\phi_* : T_p(M) \to T_{\phi(p)}(N)$, the differential map between the tangent spaces. It turns out that this also induces a map between the cotangent spaces, but in the *reverse* direction. This follows from the dual pairing between tangent and cotangent spaces.

The reverse map is denoted $\phi^* : T^*_{\phi(p)}(N) \to T^*_p(M)$ and is defined as

follows. If $\omega \in T^*_{\phi(p)}(N)$, define $\phi^*\omega \in T^*_p(M)$ by

$$\langle \phi^*\omega, X\rangle = \langle \omega, \phi_* X\rangle \qquad (X \in T_p(M)) \tag{3.51}$$

Differential forms are useful in studying the topology of general differentiable manifolds. Much of this study can be performed without introducing metric properties. It is important to keep in mind that, although a metric on a manifold is an essential construction in physics (particularly general relativity), it is an additional structure on a manifold. It is worth knowing what are all the operations one can perform without having to define a metric, since the consequences of these must clearly turn out to be metric-independent.

In the special theory of relativity, the manifold with which we deal is flat Minkowski spacetime (or Euclidean space after Wick rotation). Here differential forms play a relatively trivial role, as the topology of Euclidean space is trivial. But they do serve as useful notation, and moreover many equations of physics generalise immediately to manifolds which are different from Euclidean space, and whose topology may be nontrivial.

Example: Free electrodynamics in 4 spacetime dimensions. The electromagnetic field is represented by a 1-form $A(x) = A_\mu(x)\,dx^\mu$. The field strength is simply the 2-form $F(x) \equiv dA(x) = \frac{1}{2}F_{\mu\nu}(x)dx^\mu \wedge dx^\nu$ where

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) \tag{3.52}$$

We know that $d^2 = 0$. Since $F = dA$, it follows that $dF = d^2A = 0$. What does this condition on F physically mean? Writing it out in components, we find

$$dF = F_{\mu\nu,\lambda}\,dx^\lambda \wedge dx^\mu \wedge dx^\nu = 0 \tag{3.53}$$

By total antisymmetry of the threefold wedge product, we get

$$F_{\mu\nu,\lambda} + F_{\nu\lambda,\mu} + F_{\lambda\mu,\nu} = 0 \tag{3.54}$$

This is called the Bianchi identity.

If we temporarily get ahead of ourselves and introduce the Minkowski metric of spacetime, the above equations can be written in a more familiar form. First of all we find:

$$\begin{aligned} F_{ij,k} + \text{cyclic} = 0 \qquad & i,j,k \in \{1,2,3\} \\ F_{0i,j} + F_{ij,0} + F_{j0,i} = 0 \qquad & \end{aligned} \tag{3.55}$$

Next, defining the magnetic and electric fields by

$$B^i = \frac{1}{2}\epsilon^{ijk}F_{jk}, \quad E_i = -F_{0i} \tag{3.56}$$

the Bianchi identity becomes

$$\begin{aligned} \partial_i B^i = \vec{\nabla}\cdot\vec{\mathbf{B}} = 0 \\ \partial_i E_j - \partial_j E_i = -\epsilon_{ijk}\frac{\partial B^k}{\partial t} \rightarrow \vec{\nabla}\times\vec{\mathbf{E}} = -\frac{\partial \vec{B}}{\partial t} \end{aligned} \tag{3.57}$$

So the Bianchi identity, $d^2A = dF = 0$, is just equivalent to two of Maxwell's equations in empty space.

It is tempting at this point to display the two remaining Maxwell equations in the same notation. For this, we must first define the constant antisymmetric tensor (under $SO(3,1)$ Lorentz transformations) on $\mathbb{R}^4$: it is denoted $\epsilon_{\mu\nu\lambda\rho}$, and takes the value $+1$ if $\{\mu,\nu,\lambda,\rho\}$ is an even permutation of $\{0,1,2,3\}$ and -1 if it is an odd permutation. It is equal to 0 if any two of the indices coincide. We can think of this tensor as making up the components of the 4-form

$$\lambda = \frac{1}{4!}\epsilon_{\mu\nu\lambda\rho} dx^\mu \wedge dx^\nu \wedge dx^\lambda \wedge dx^\rho \tag{3.58}$$

We can use $\epsilon_{\mu\nu\lambda\rho}$ to define what we call the *dual* of any arbitrary form. (We will discuss this concept in more generality in subsequent chapters). For a 1-form $a = a_\mu\, dx^\mu$, we define its *dual 3-form*

$$^*a = \frac{1}{3!}{}^*a_{\mu\nu\lambda}\, dx^\mu \wedge dx^\nu \wedge dx^\lambda \tag{3.59}$$

by specifying the components as $a_{\mu\nu\lambda} = \epsilon_{\mu\nu\lambda\rho}\, a^\rho$.

Similarly, the dual of a 2-form $b = \frac{1}{2}b_{\mu\nu}\, dx^\mu \wedge dx^\nu$ is the 2-form $^*b = \frac{1}{2}{}^*b_{\mu\nu}\, dx^\mu \wedge dx^\nu$ with components

$$^*b_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\lambda\rho}\, b^{\lambda\rho} \tag{3.60}$$

It is easy to see similarly that in general, the dual of an n-form is a $(4-n)$-form, and that taking the dual of any form twice gives the original form back again: $^*(^*a) = a$ for any n-form a.

Now consider $^*F$ where F is the Maxwell field strength 2-form defined above. Clearly, in general we do not have $d^*F = 0$, since $^*F$ (unlike F itself) is not d of anything. Nevertheless, let us examine the content of the equation $d(^*F) = 0$, which could be imposed by physical rather than mathematical requirements. The resulting equations, in component form, will turn out to be the other two Maxwell equations. Thus, these equations will be seen to have a *dynamical* content, and they have to be imposed by hand, unlike the two we already obtained, which are just mathematical identities.

Working this out in components we have:

$$d(^*F) = \frac{1}{2}\epsilon_{\mu\nu\lambda\rho}\, \partial_\alpha F^{\lambda\rho}\, dx^\sigma \wedge dx^\mu \wedge dx^\nu \tag{3.61}$$

Setting this to zero, and taking the dual of the equation we find:

$$\epsilon^{\sigma\mu\nu\alpha}\, \epsilon_{\nu\nu\lambda\rho}\, \partial_\sigma F^{\lambda\rho}\, dx^\alpha = 0 \tag{3.62}$$

which after contracting the indices of the two ϵ-symbols gives

$$\partial^\mu F_{\mu\nu} = 0 \tag{3.63}$$

These are just the following two empty-space Maxwell equations:

$$\vec{\nabla} \cdot \vec{E} = 0, \qquad \vec{\nabla} \times \vec{B} = \frac{\partial \vec{E}}{\partial t} \tag{3.64}$$

Thus in differential form language, the free Maxwell's equations are summarized in:

$$\begin{aligned} dF &= 0 \qquad && \text{(identity)} \\ d^*F &= 0 && \text{(dynamical equation)} \end{aligned} \tag{3.65}$$

In realistic physical situations the second equation acquires a term on the right hand side representing an electrical charge or current source.

Finally, note that the fundamental concept of *gauge invariance* in Maxwell's equations has a simple interpretation in terms of differential forms. If we change the vector potential by $A \to A' = A + d\Lambda$ (where Λ is a function, or 0-form), then

$$F \to F' = dA' = d(A + d\Lambda) = dA = F \tag{3.66}$$

because $d^2 = 0$. This is gauge invariance for free electrodynamics.

3.7 More About Vectors and Forms

Here we will display some more properties of the structures defined above. In particular, this will lead to a binary operation, the "Lie bracket", which maps a pair of vector fields to a new vector field.

Choose a vector field $X = a^i(x)\,\partial/\partial x^i$, and a 1-form $\omega = \omega_i(x)\,dx^i$ Recall that X acts on a function f by $X : f \to Xf$ to give a new function $a^i(x)\,\frac{\partial f}{\partial x^i}$. Also, the inner product between vector fields and 1-forms is a function:

$$\langle \omega, X \rangle = \omega_i(x)\, a^i(x) \tag{3.67}$$

Now given two vector fields, we can act with them successively on a function, in two different orders, and find the commutator of the two actions. Thus, if $X = a^i(x)\,\partial/\partial x^i$, $Y = b^i(x)\,\partial/\partial x^i$, consider

$$f \to Yf \to XYf = a^i \frac{\partial}{\partial x^i}\left(b^j \frac{\partial f}{\partial x^j}\right) \tag{3.68}$$

and,

$$f \to Xf \to YXf = b^i \frac{\partial}{\partial x^i}\left(a^j \frac{\partial f}{\partial x^j}\right) \tag{3.69}$$

Computing the difference of these two operations:

$$XYf - YXf = \left(a^i \frac{\partial b^j}{\partial x^i} - b^i \frac{\partial a^j}{\partial x^i}\right) \frac{\partial f}{\partial x^j} \tag{3.70}$$

This defines a new vector field:

Definition: The *Lie bracket* of two vector fields X and Y is the vector field

$$[X,Y]: \quad f \to [X,Y]f = XYf - YXf \tag{3.71}$$

$[X,Y]$ is sometimes also denoted $L_X Y$, the *Lie derivative* of Y along X. In components,

$$[X,Y] = \left(a^i \frac{\partial b^j}{\partial x^i} - b^i \frac{\partial a^j}{\partial x^i}\right) \frac{\partial}{\partial x^j} \tag{3.72}$$

The Lie bracket gives a Lie algebra structure to the tangent space. This arises from the the following (easily derived) identities:

$$\begin{aligned} [X,Y] &= -[Y,X] \\ [X_1 + X_2, Y] &= [X_1,Y] + [X_2,Y] \\ [cX,Y] &= c[X,Y], \qquad c \in \mathbb{R} \end{aligned} \tag{3.73}$$

as well as the Jacobi identity:

$$[[X,Y],Z] + [[Y,Z],X] + [Z,[X,Y]] = 0 \tag{3.74}$$

Next, we prove an interesting identity about 2-forms. Since a 1-form maps tangent vector fields to real functions,

$$\omega: \quad X \to \langle \omega, X \rangle, \tag{3.75}$$

a 2-form will do the same for a *pair* of vectors:

$$\Omega = \Omega_{ij}(x)\, dx^i \wedge dx^j$$

$$\Omega: \quad (X,Y) \to \langle \Omega; X, Y \rangle \equiv \Omega_{ij}(x)\, a^i(x)\, b^j(x) \tag{3.76}$$

Now if we are given a 1-form ω and two vector fields X and Y:

$$\omega = \omega_i(x)\, dx^i, \quad X = a^i(x) \frac{\partial}{\partial x^i}, \quad Y = b^i(x) \frac{\partial}{\partial x^i} \tag{3.77}$$

then $d\omega$ is a 2-form and we have the identity:

$$\langle d\omega; X, Y \rangle = \frac{1}{2}\{X\langle \omega, Y\rangle - Y\langle \omega, X\rangle - \langle \omega, [X,Y]\rangle\} \tag{3.78}$$

The proof follows by expanding both sides:

$$\begin{aligned} \text{LHS} &= \frac{1}{2}(\partial_i \omega_j - \partial_j \omega_i)\, a^i\, b^j \\ \text{R.H.S.} &= \frac{1}{2}\left[a^i \frac{\partial}{\partial x^i}(\omega_j b^j) - b^i \frac{\partial}{\partial x^i}(\omega_j a^j) - \omega_i \left(a^j \frac{\partial b^i}{\partial x^j} - b^j \frac{\partial a^i}{\partial x^j}\right)\right] \\ &= \frac{1}{2}(\partial_i \omega_j - \partial_j \omega_i)\, a^i\, b^j \end{aligned} \tag{3.79}$$

The Lie bracket of vector fields appears in the last term in this identity, and indeed the identity can be turned around to provide an alternative definition of the Lie bracket.

Chapter 4

Differentiable Manifolds II

4.1 Riemannian Geometry

So far we have studied differentiable manifolds without assigning any metric to them. The properties of spaces that we have discussed so far – continuity for topological spaces, and differentiability for manifolds – do not require the assignment of a metric. However, manifolds equipped with a notion of distance are fundamental in physics, so we now turn to the study of Riemannian manifolds, namely manifolds with a metric.

Since a manifold is only locally like Euclidean space, we will have to start by working locally. Moreover we must make sure that whatever distance we define between two points will not depend on the specific coordinate systems around the points.

On $\mathbb{R}^n$ with Cartesian coordinates[1] x^μ, there is a natural notion of distance between two infinitesimally separated points x^μ and $x^\mu + \delta x^\mu$ given by

$$ds^2 = \delta x^\mu \delta x^\mu$$

We can use this fact to define a distance on a general differentiable manifold M. Here, in a given chart with coordinates x^μ, consider two points p, q whose coordinates differ infinitesimally:

$$p \to x^\mu, \; q \to x^\mu + \delta x^\mu$$

Suppose another chart with coordinates y^μ overlaps with the first one. Then on the overlap of charts, the displacement δx^μ is given in terms of δy^μ by the chain rule:

$$\delta x^\mu = \frac{\partial x^\mu}{\partial y^\nu} \delta y^\nu \tag{4.1}$$

[1] In this section we use $\mu = 1, 2, \cdots, d$ to label the directions of a d-dimensional manifold. This notation is common in physics. Summation over repeated indices is implied, as before.

Clearly

$$\delta x^\mu \delta x^\mu = \frac{\partial x^\mu}{\partial y^\lambda}\frac{\partial x^\mu}{\partial y^\rho}\,\delta y^\lambda \delta y^\rho \neq \delta y^\mu \delta y^\mu \tag{4.2}$$

so the Cartesian definition of distance is not coordinate-independent.

However, the above equation tells us how to modify the definition such that it becomes coordinate independent. We need to "compensate" the transformation of coordinates by introducing a quantity that transforms in the opposite way. Therefore we define a rank-2 covariant tensor field $g_{\mu\nu}(x)$:

$$g_{\mu\nu}(x) \in T_p^*(M) \otimes T_p^*(M) \tag{4.3}$$

where x is the coordinate of the point p. Given such a tensor, the distance $d(p,q)$ between points p and q (infinitesimally separated in terms of their coordinates in the given chart) is defined by:

$$[d(p,q)]^2 \equiv ds^2 = g_{\mu\nu}(x)\,\delta x^\mu \delta x^\nu \tag{4.4}$$

This is coordinate independent by virtue of the transformation law of a rank-2 covariant tensor field:

$$g'_{\mu\nu}(y) = \frac{\partial x^\lambda}{\partial y^\mu}\frac{\partial x^\rho}{\partial y^\nu}\,g_{\lambda\rho}(x) \tag{4.5}$$

from which, using the chain rule, it is easy to see that:

$$g'_{\mu\nu}(y)\,\delta y^\mu \delta y^\nu = g_{\mu\nu}(x)\,\delta x^\mu \delta x^\nu \tag{4.6}$$

The tensor $g_{\mu\nu}$ must be chosen so that the three axioms defined in Chapter 1 for a metric on a topological space are satisfied. $g_{\mu\nu}$ is called the *Riemann metric* or simply *metric tensor*.

The above definition only tells us the distance between infinitesimally separated points. To extend it to compute the total length of any path between *finitely* separated points, simply integrate the infinitesimal ds defined above along the given path. We will define the distance between finitely-separated points p, q to be the *minimum* of all path lengths between the points:

$$d(p,q) = \min \int_p^q ds \tag{4.7}$$

The integral is evaluated on the image of the path in some chosen chart.

We may write a Riemannian metric as $G = g_{\mu\nu}(x)\,dx^\mu \otimes dx^\nu$ where the notation $\otimes$ highlights the fact that it is similar to a differential form except that it is not antisymmetric. Such an object can be contracted with two vector fields using the duality of tangent and cotangent spaces that we have already noted:

$$\langle dx^\mu, \frac{\partial}{\partial x^\nu}\rangle = \delta^\mu_{\ \nu} \tag{4.8}$$

Thus, if $A = a^\mu(x)\frac{\partial}{\partial x^\mu}$ and $B = b^\mu(x)\frac{\partial}{\partial x^\mu}$ are two vector fields, then

$$\begin{aligned}\langle G; A, B\rangle &= \langle\, g_{\mu\nu}(x)\,dx^\mu \otimes dx^\nu;\ a^\lambda(x)\frac{\partial}{\partial x^\lambda},\ b^\rho(x)\frac{\partial}{\partial x^\rho}\rangle \\ &= g_{\mu\nu}(x)\,a^\mu(x) b^\nu(x)\end{aligned} \tag{4.9}$$

Thus, the metric defines an *inner product* on the space of vector fields:

$$\langle A, B\rangle_G \equiv \langle G; A, B\rangle = g_{\mu\nu}\, a^\mu b^\nu \tag{4.10}$$

This provides new insight into the meaning of the Riemannian metric.

Indeed, there is a relation between the two roles of the metric: that of providing a distance on the manifold, and that of providing an inner product on vector fields. To see this, note that:

$$\begin{aligned} d(p,q) &= \min \int_p^q ds = \min \int_p^q \sqrt{g_{\mu\nu}(x) dx^\mu dx^\nu} \\ &= \min \int_0^1 \sqrt{g_{\mu\nu} \frac{dx^\mu}{dt} \frac{dx^\nu}{dt}}\, dt \end{aligned} \tag{4.11}$$

where we have parametrised the image of each curve joining p and q as $x^\mu(t)$ with $0 \le t \le 1$, with $x^\mu(0), x^\mu(1)$ being the coordinates of p, q respectively.

Now $\frac{dx^\mu}{dt}$ are just the components of the tangent vector:

$$T = \frac{dx^\mu}{dt} \frac{\partial}{\partial x^\mu} \tag{4.12}$$

to the curve. So

$$d(p,q) = \min \int_0^1 \sqrt{\langle T, T\rangle_G}\, dt \tag{4.13}$$

This expresses the distance between two points along a given curve as an integral over the norm of the tangent vector to the curve.

Definition: A differentiable manifold equipped with a Riemannian metric as defined above is said to be a *Riemannian manifold.*

4.2 Frames

It is useful to define a *basis* for the tangent space at each point of a manifold. A continuously varying basis will correspond to a collection of d vector fields on M, where n is the dimension of M.

We can choose the set to be orthonormal in the inner product on $T_p(M)$ defined by the metric. If the vector fields are denoted

$$E_a(x) = E_a^\mu(x) \frac{\partial}{\partial x^\mu}, \quad a = 1, \cdots, d \tag{4.14}$$

then this amounts to requiring that:

$$\langle E_a, E_b\rangle_G \equiv g_{\mu\nu}(x) E_a^\mu(x) E_b^\nu(x) = \delta_{ab} \tag{4.15}$$

Definition: Vector fields $E_a(x)$ satisfying the above requirements are called *orthonormal frames* or *vielbeins*. Typically they are just referred to as "frames".

We may also define the 1-forms dual to these vector fields, These are denoted:

$$e^a(x) = e^a_\mu(x)\, dx^\mu \tag{4.16}$$

and are defined to satisfy:

$$\langle e^a, E_b \rangle = e^a_\mu(x)\, E^\mu_b(x) = \delta^a{}_b \tag{4.17}$$

For a d-dimensional manifold, an $O(d)$ rotation of the frames gives a new set of orthonormal frames (here $O(d)$ denotes the group of $d \times d$ orthogonal matrices). The duality between e^a and E_a is preserved if the same $O(d)$ rotation is made on both.

The frames and their dual 1-forms in turn determine the metric tensor of the manifold. From the 1-form e^a we can construct the $O(d)$-invariant symmetric second-rank tensor, $\sum_{a=1}^d e^a \otimes e^a$. As we have seen, such a tensor defines an inner product on tangent vectors. Thus we may compute:

$$\begin{aligned} \langle \sum_{a=1}^{d} e^a \otimes e^a; E_b, E_c \rangle &= e^a_\mu\, e^a_\nu\, E^\mu_b\, E^\nu_c \\ &= \delta_{bc} \\ &= \langle E_b, E_c \rangle_G \end{aligned} \tag{4.18}$$

from which we conclude that

$$G \equiv g_{\mu\nu} dx^\mu \otimes dx^\nu = \sum_{a=1}^{n} e^a \otimes e^a \tag{4.19}$$

or in component notation,

$$g_{\mu\nu}(x) = e^a_\mu(x) e^a_\nu(x). \tag{4.20}$$

(the repeated index is summed over).

Similarly, if we define a rank-2 tensor field $\sum_{a=1}^n E_a \otimes E_a$, then we can write

$$\sum E_a \otimes E_a = h^{\mu\nu} \frac{\partial}{\partial x^\mu} \otimes \frac{\partial}{\partial x^\nu} \tag{4.21}$$

from which it follows easily that

$$h^{\mu\nu}(x) g_{\nu\lambda}(x) = \delta^\mu{}_\lambda \tag{4.22}$$

So $h^{\mu\nu}$ is the matrix inverse of $g_{\mu\nu}$. Henceforth we write $h^{\mu\nu}$ as $g^{\mu\nu}$.

Exercise: Check all the manipulations above.

Once we are equipped with a set of orthonormal frames and their dual 1-forms, it becomes convenient to study vectors and tensors by referring their

components to this basis. For example, given any vector field $A(x) = A^\mu(x)\frac{\partial}{\partial x^\mu}$, we can take its inner product with the 1-form $e^a(x)$:

$$A^a(x) \equiv \langle e^a(x), A(x)\rangle = e^a_\mu(x)A^\mu(x) \tag{4.23}$$

The n-component objects $A^a(x)$ are a collection of (scalar) functions on the manifold. Under changes of the coordinate system, both $e^a_\mu(x)$ and $A^\mu(x)$ change such that their inner product remains invariant.

Similarly, a 1-form $B(x) = B_\mu(x)dx^\mu$ can be converted into a collection of (scalar) functions:

$$B_a(x) = \langle B(x), E_a(x)\rangle = E^\mu_a(x)B_\mu(x) \tag{4.24}$$

Although coordinate invariant, the functions $A^a(x)$ and $B_a(x)$ do depend on the choice of orthonormal frames, and they change under $O(n)$ rotations of these frames, for example:

$$B_a(x) \to \Lambda_a{}^b(x)B_b(x), \quad \Lambda^T\Lambda = 1 \tag{4.25}$$

So it may seem we have not gained anything by converting the form into $O(n)$ vector-valued functions. Indeed, going from forms to $O(n)$ vectors is completely reversible and this is also true if we extend the above relation to map coordinate tensors and $O(n)$ tensors to each other via:

$$\begin{aligned} B_{a_1a_2\cdots a_n} &= E^{\mu_1}_{a_1}E^{\mu_2}_{a_2}\cdots E^{\mu_n}_{a_n}B_{\mu_1\mu_2\cdots\mu_n} \\ B_{\mu_1\mu_2\cdots\mu_n} &= e^{a_1}_{\mu_1}e^{a_2}_{\mu_2}\cdots e^{a_n}_{\mu_n}B_{a_1a_2\cdots a_n} \end{aligned} \tag{4.26}$$

This suggests that it is merely a matter of convenience to work with quantities that transform under coordinate transformations (tensor fields and forms) as against those that transform under $O(n)$.

However the utility of orthonormal frames goes beyond convenience. $O(n)$ is a semi-simple Lie group and its representations are well-understood. In particular (see N. Mukunda's lectures in this volume) it admits a family of *spinor* representations. Fields that transform in these representations cannot be mapped to (or from) coordinate tensors or differential forms as above. Therefore by introducing frames and using fields that transform in $O(n)$ representations, we are able to deal with a more general class of objects on manifolds. This possibility is of crucial importance in the physical context of general relativity, since this is precisely how *fermions* are introduced.

4.3 Connections, Curvature and Torsion

Recall that we defined the exterior derivative on 1-forms as[2]

$$d\colon\ A = A_\mu\, dx^\mu \to dA = \partial_\nu A_\mu\, dx^\nu \wedge dx^\mu \tag{4.27}$$

[2]As usual the components of the forms depend on coordinates, but from this section onwards we suppress the argument to simplify our notation.

We have shown that dA obtained in this way is indeed a 2-form.

Now suppose we first convert the 1-form A into a zero-form using the frames, as in the previous section, and then take the exterior derivative. Thus we start by defining:

$$A_a = \langle A, E_a \rangle = E_a^\mu A_\mu \tag{4.28}$$

and then attempt to define:

$$dA_a = \partial_\mu A_a \, dx^\mu \tag{4.29}$$

Is the result sensible? It is certainly a differential form. But it is easy to see that it fails to be an $O(n)$ vector. In fact, under a local $O(n)$ rotation of the frames,

$$A_a \to A'_a = \Lambda_a{}^b A_b, \quad \Lambda_a{}^b \in O(n) \tag{4.30}$$

one finds that:

$$dA_a \to dA'_a = d\Lambda_a{}^b A_b + \Lambda_a{}^b dA_b \tag{4.31}$$

Because of the first term, which involves a derivative of $\Lambda_a{}^b$, this is not how an $O(n)$ vector transforms.

Therefore we need to look for a new type of derivative D, called a *covariant derivative*, on $O(n)$ vectors. It is required to have the property that, when acting on $O(n)$ vectors, it gives us back $O(n)$ vectors. To qualify as a derivative, D must be a linear operation, so it is natural to try a definition like:

$$(DA)_a = dA_a + \omega_a{}^b A_b \tag{4.32}$$

Here,

$$\omega_a{}^b = \omega_{\mu\, a}{}^b \, dx^\mu \tag{4.33}$$

is a 1-form whose $O(n)$ transformation rules remain to be found, and will be determined by requiring that the result of differentiation is again an $O(n)$ vector.

For this, we simultaneously carry out the known $O(n)$ transformation on the 1-form and an arbitrary transformation on ω:

$$\begin{aligned} A_a &\to A'_a = \Lambda_a{}^b A_b \\ \omega_a{}^b &\to \omega'_a{}^b \end{aligned} \tag{4.34}$$

Then

$$\begin{aligned} (DA')_a &= dA'_a + \omega'_a{}^b A'_b \\ &= d\Lambda_a{}^b A_b + \Lambda_a{}^b dA_b + \omega'_a{}^b \Lambda_b{}^c A_c \\ &= \Lambda_a{}^b \Big(dA_b + (\Lambda^{-1})_b{}^c d\Lambda_c{}^d A_d + (\Lambda^{-1})_b{}^c \omega'_c{}^d \Lambda_d{}^e A_e \Big) \end{aligned} \tag{4.35}$$

Requiring $(DA')_a = \Lambda_a{}^b \, DA_b$ as expected of an $O(n)$ vector, and comparing terms, we find:

$$\Lambda^{-1} d\Lambda + \Lambda^{-1} \omega' \Lambda = \omega \tag{4.36}$$

where matrix multiplication is intended. Hence

$$\omega' = \Lambda\,\omega\Lambda^{-1} - d\Lambda\,\Lambda^{-1} \tag{4.37}$$

Thus we have shown that $\omega_a{}^b$ transforms *inhomogeneously* under $O(n)$. It is called a *connection* (sometimes, in this context, the *spin connection*).

We can extract an important property of the spin connection from the above transformation law. If we take the transpose and use the fact that $\Lambda^T = \Lambda^{-1}$, we find:

$$\omega'^T = \Lambda\,\omega^T\Lambda^{-1} + d\Lambda\,\Lambda^{-1} \tag{4.38}$$

Adding and subtracting the two equations above, we see that the symmetric part of ω transforms as a tensor, while the antisymmetric part has the characteristic inhomogeneous transformation law[3].

Antisymmetry of ω, as demonstrated above, allows us to check that the following identity holds:

$$d(A_a\,A^a) = (DA)_a\,A^a + A_a\,(DA)^a \tag{4.39}$$

On general $O(n)$ tensors, the definition of the covariant derivative naturally extends to:

$$(DA)_{ab\cdots c} = dA_{ab\cdots c} + \omega_a{}^d A_{ab\cdots c} + \omega_b{}^d A_{ad\cdots c} + \cdots + \omega_c{}^d A_{ab\cdots d} \tag{4.40}$$

Exercise: Check that the $O(n)$ transformation $A_{ab\cdots c} \to \Lambda_a{}^p\Lambda_b{}^q\cdots\Lambda_c{}^r A_{pq\cdots r}$ together with the transformation for ω discovered above, transforms the covariant derivative $(DA)_{ab\cdots c}$ in the same way as A.

The spin connection is a differential 1-form but is not itself an $O(n)$ tensor, as is evident from the inhomogeneous transformation law written above. However, using the spin connection we can define two very basic tensors associated to a manifold. Applying D to the orthonormal frames e^a (which are $O(n)$ vector-falued 1-forms), one gets an $O(n)$ vector-valued 2-form.

Definition: The 2-form:

$$T^d \equiv De^a = de^a + \omega^a{}_b \wedge e^b \tag{4.41}$$

is called the *torsion 2-form* on the manifold M.

Another tensor arises naturally by asking the following question. We have seen that in the absence of a spin connection, the exterior derivative d satisfies

[3]Note that it makes no difference whether the $O(n)$ indices are raised or lowered, as this is done using the identity metric δ_{ab}. In applications to general relativity the group $O(n)$ that acts on frames will be replaced by the Lorentz group $O(n-1,1)$ in which case the metric that raises and lowers these indices is the Minkowski metric $\eta_{ab} = \mathrm{diag}(-1,1,\cdots,1)$. Here too the difference between raised and lowered indices is trivial, involving at most a change of sign.

$d^2 = 0$. But what about D^2, the square of the covariant exterior derivative D? This is easily calculated by acting twice with D on an arbitrary $O(n)$ vector A^a:

$$\begin{aligned}(D^2A)^a = \Big(D(DA)\Big)^a &= d(dA^a + \omega^a{}_b A^b) + \omega^a{}_c \wedge (dA^c + \omega^c{}_b A^b) \\ &= d\omega^a{}_b A^b - \omega^a{}_b \wedge dA^b + \omega^a{}_b \wedge dA^b + \omega^a{}_c \wedge \omega^c{}_b A^b \\ &= (d\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b) A^b \equiv R^a{}_b A^b \end{aligned} \tag{4.42}$$

Definition: The *curvature 2-form* is defined by:

$$R^a{}_b \equiv d\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b \tag{4.43}$$

which in components may be represented:

$$\begin{aligned} R^a{}_b &= \frac{1}{2} R^a{}_{b\,\mu\nu}\, dx^\mu \wedge dx^\nu \\ &= \frac{1}{2} R^a{}_{bcd}\, e^c \wedge e^d \end{aligned} \tag{4.44}$$

The torsion and curvature 2-forms are fundamental to the study of Riemannian manifolds.

Exercise: Check the following properties of T^a and R^a_b:

$$\begin{aligned} (DT)^a &\equiv dT^a + \omega^a{}_b \wedge T^b = R^a{}_b \wedge e^b \\ (DR)^a{}_b &\equiv dR^a{}_b + \omega^a{}_c \wedge R^c{}_b + \omega^c{}_b \wedge R^a{}_c = 0 \end{aligned} \tag{4.45}$$

The second of these relations is called the *Bianchi identity.*

Under $O(n)$ rotations, the torsion and curvature transform respectively as

$$\begin{aligned} T^a &\to \Lambda^a{}_b T^b \\ R^a_b &\to \Lambda^a{}_c \Lambda_b{}^d R^c{}_d = (\Lambda R \Lambda^{-1})^a{}_b \end{aligned} \tag{4.46}$$

Let us now turn to a new type of connection. We have seen that a spin connection is needed in order to differentiate $O(n)$ tensors covariantly. Suppose instead that we wish to covariantly differentiate a *vector field*

$$A = A^\mu \frac{\partial}{\partial x^\mu} \tag{4.47}$$

As before, the ordinary derivative produces something that does not transform as a tensor:

$$dA = \partial_\nu A^\mu\, dx^\nu \otimes \frac{\partial}{\partial x^\mu} \tag{4.48}$$

So again, we postulate a covariant derivative:

$$DA = D_\nu A^\mu dx^\nu \otimes \frac{\partial}{\partial x^\mu} \tag{4.49}$$

such that $D_\nu A^\mu$ transforms as a tensor under changes of coordinates. Requiring linearity, we assume

$$D_\nu A^\mu \equiv \partial_\nu A^\mu + \Gamma^\mu_{\nu\lambda} A^\lambda \tag{4.50}$$

The quantity $\Gamma^\mu_{\nu\lambda}$ is called the *affine connection*. In order for DA to transform covariantly under general coordinate transformations $x^\mu \to x'^\mu(x)$, the affine connection must transform as

$$\Gamma'^\mu_{\nu\lambda}(x') = \frac{\partial x'^\mu}{\partial x^\alpha}\frac{\partial x^\beta}{\partial x'^\nu}\frac{\partial x^\nu}{\partial x'^\lambda}\Gamma^\alpha_{\beta\nu}(x) + \frac{\partial x'^\mu}{\partial x^\alpha}\frac{\partial^2 x^\alpha}{\partial x'^\nu \partial x'^\lambda} \tag{4.51}$$

Exercise: Check that with the above transformation law, the covariant derivative of a vector field transforms covariantly under general coordinate transformations.

On 1-forms $B = B_\mu dx^\mu$, the covariant derivative is defined similarly, but with an important change of sign and different index contractions:

$$DB = D_\nu B_\mu dx^\nu \otimes dx^\mu$$
$$D_\nu B_\mu \equiv \partial_\nu B_\mu - \Gamma^\lambda_{\mu\nu} B_\lambda \tag{4.52}$$

This is required so that the contraction of a covariant and contravariant tensor behaves as a scalar.

Note that DB as we have defined it above is not an antisymmetric 2-form, but just a general rank-2 covariant tensor. We could, of course, take the antisymmetric part of it to get a 2-form. In that situation, recalling the properties of exterior derivatives, one would not need any connection at all. From the above equation, we find that the antisymmetrised D acting on a 1-form contains as a connection the object $\Gamma^\lambda_{\mu\nu} - \Gamma^\lambda_{\nu\mu}$, which is called the *torsion* associated to the affine connection[4]. It is easy to check that this torsion transforms as a tensor, and can be chosen to vanish consistently with the covariance properties of the derivative.

Returning now to the spin connection ω, we have observed above that its symmetric part transforms as a tensor. Thus it is not really required in order to fulfill the main objective of a connection, which is to provide a rule for covariant differentiation. Therefore one (standard) way to specify ω is to require that it be antisymmetric. If in addition we require the torsion (the covariant derivative of ω) to vanish, it can easily be seen that ω is completely determined in terms of the frames e^a. To see this, let us impose:

$$\begin{aligned} \text{antisymmetry:} &\quad \omega_{ab} = -\omega_{ba} \\ \text{no torsion:} &\quad T^a \equiv (De)^a = 0 \end{aligned} \tag{4.53}$$

Exercise: Show that the above conditions imply the relation:

$$\omega^a_{\mu\, b} = \frac{1}{2}E^{a\nu}\left[(\partial_\mu e^b_\nu - \partial_\nu e^b_\mu) - e^c_\mu E^{b\sigma}(\partial_\nu e_{c\sigma} - \partial_\sigma e_{c\nu})\right] - (a \leftrightarrow b) \tag{4.54}$$

[4]This is distinct from the torsion defined above in terms of orthonormal frames.

The ω so defined is sometimes known as the *Levi-Civita spin connection.*

Similarly, an affine connection Γ can be uniquely specified by two conditions:

$$\begin{aligned} \text{metricity}: &\quad D_\mu g_{\nu\lambda} \equiv \partial_\mu g_{\nu\lambda} - \Gamma^\alpha_{\mu\nu} g_{\alpha\lambda} - \Gamma^\alpha_{\mu\lambda} g_{\nu\alpha} = 0 \\ \text{no torsion}: &\quad \Gamma^\mu_{\nu\lambda} - \Gamma^\mu_{\lambda\nu} = 0 \end{aligned} \tag{4.55}$$

Metricity is the same as covariant constancy of the metric.

The affine connection is uniquely determined in terms of the metric by the two conditions above:

$$\Gamma^\mu_{\nu\lambda} = \frac{1}{2} g^{\mu\alpha} (g_{\alpha\nu,\lambda} + g_{\alpha\lambda,\nu} - g_{\nu\lambda,\alpha}) \tag{4.56}$$

This is called the *Levi-Civita affine connection*, or *Christoffel symbol.*

Exercise: Demonstrate that Eq.(4.56) follows from the two conditions in Eq.(4.55) above.

The spin connection and affine connection can be related to each other by requiring covariant constancy of the frames:

$$D_\mu e^a_\nu = \partial_\mu e^a_\nu + \omega^a_{\mu b} e^b_\nu - \Gamma^\lambda_{\mu\nu} e^a_\lambda = 0 \tag{4.57}$$

This equation clearly determines either connection in terms of the other one.

Exercise: Solve the above equation for the spin connection in terms of the frames and the affine connection. Next, use Eq.(4.56) to express the affine connection in terms of the metric and thence in terms of frames using $g_{\mu\nu} = e^a_\mu e^a_\nu$. At the end, you will have an expression for the spin connection in terms of the frames. Check that this is identical to Eq.(4.54).

From all the above, it is evident that both the spin connection and the affine connection are essential ingredients in any system where we would like to differentiate vector/tensor fields, differential forms, and $O(n)$ tensors on a manifold. While there is a certain degree of arbitrariness in their definition, there are certain "minimal" conditions which can be imposed to render them unique. In applications to physics, it will turn out that these conditions are naturally satisfied in the theories of interest. The corresponding connections are then *dependent* variables, being determined by the frames in the case of the spin connection and by the metric in the case of the affine connection[5].

4.4 The Volume Form

Recall that the group $SO(n)$ differs from the orthogonal group $O(n)$ in that in the former, the rotation matrices are required to have unit determinant in

[5]Some physical applications appear to require an affine connection that has nonzero torsion (i.e. is not symmetric in its lower indices). However, since the torsion so defined is a tensor, it can always be treated separately from the connection and this choice usually proves more convenient in practice.

addition to being orthogonal. Thus, reflection of an odd number of space dimensions, which constitutes an orthogonal transformation disconnected from the identity, is not included in $SO(n)$ though it is part of $O(n)$.

Now for an *orientable* n-dimensional manifold we should ignore those frame rotations that are not continuously connected to the identity, as such rotations would reverse the orientation. Therefore we restrict ourselves to $SO(n)$ rotations of the frames rather than $O(n)$. Indeed, frames of a given *handedness* can be chosen continuously everywhere on an orientable manifold.

Consider the n-form:

$$\lambda = e^1 \wedge e^2 \wedge \cdots \wedge e^n \tag{4.58}$$

This is invariant under $SO(n)$ rotations, as the following exercise shows.

Exercise: Prove that when $e^a \to \Lambda^a{}_b e^b$ with $\Lambda^a{}_b \in O(n)$, λ changes to $(\det \Lambda)\lambda$. Hence if $\det \Lambda = 1$, (equivalently, $\Lambda \in SO(n)$) then λ is invariant.

Definition: The n-form λ defined above is called a *Riemannian volume form.*

We will see that the volume form plays a crucial role in integration theory on manifolds.

We can rewrite the volume form as

$$\begin{aligned}\lambda &= e^1_{\mu_1} e^2_{\mu_2} \cdots e^n_{\mu_n} dx^{\mu_1} \wedge dx^{\mu_2} \wedge \ldots \wedge dx^{\mu_n} \\ &= (\det e)\, dx^1 \wedge dx^2 \wedge \cdots \wedge dx^n \\ &= \sqrt{g}\; dx^1 \wedge dx^2 \wedge \cdots \wedge dx^n \end{aligned} \tag{4.59}$$

where $g \equiv \det g_{\mu\nu}$.

Another useful way to write it is in terms of the totally antisymmetric ϵ-tensor, defined as:

$$\begin{aligned}\epsilon_{\mu_1 \cdots \mu_n} &= 0 \quad \text{(if two indices coincide)} \\ &= +1 \quad \text{(if } (\mu_1 \cdots \mu_n) \text{ is an even permutation)} \\ &= -1 \quad \text{(if } (\mu_1 \cdots \mu_n) \text{ is an odd permutation)} \end{aligned} \tag{4.60}$$

(An even (odd) permutation means that $(\mu_1 \cdots \mu_n)$ is obtained from $(1 \cdots n)$ by an even (odd) number of pairwise interchanges.) In terms of this tensor, the volume form may be written

$$\lambda = \frac{\sqrt{g}}{n!}\, \epsilon_{\mu_1 \cdots \mu_n} dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_n}$$

One important role of the volume form is that it allows us to define a "duality" operation which maps p-forms to $(n-p)$-forms. This is carried out via the *Hodge dual* operation $*$, defined by:

$$*(dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_p}) = \frac{\sqrt{g}}{(n-p)!}\, \epsilon^{\mu_1 \cdots \mu_p}{}_{\mu_{p+1} \cdots \mu_n}\, dx^{\mu_{p+1}} \wedge \ldots \wedge dx^{\mu_n}$$

Equivalently, if A is a p-form:

$$A = a_{\mu_1 \cdots \mu_p} dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_p}$$

then $^*A$ is the $(n-p)$ form satisfying

$$A \wedge {}^*A = g^{\mu_1 \nu_1} g^{\mu_2 \nu_2} \cdots g^{\mu_p \nu_p} a_{\mu_1 \cdots \mu_p} a_{\nu_1 \cdots \nu_p} \lambda$$

where λ is the volume form. It follows that

$${}^*A = \frac{\sqrt{g}}{(n-p)!} a_{\mu_1 \cdots \mu_p} \epsilon^{\mu_1 \cdots \mu_p}{}_{\mu_{p+1} \cdots \mu_n} dx^{\mu_{p+1}} \wedge \cdots \wedge dx^{\mu_n}$$

Since the above discussion is somewhat abstract, let us give a couple of concrete examples. In three dimensions with the Euclidean (identity) metric, the volume form is:

$$\lambda = dx^1 \wedge dx^2 \wedge dx^3 \tag{4.61}$$

The Hodge dual of a 1-form is a 2-form via:

$${}^*dx^1 = dx^2 \wedge dx^3 \tag{4.62}$$

The familiar "cross product" in vector analysis makes use of this dual. Given two 3-vectors $\vec{v}$ and $\vec{w}$, one defines two 1-forms:

$$v = v_i \, dx^i, \qquad w = w_j \, dx^j \tag{4.63}$$

The wedge product of these vectors is the 2-form:

$$v \wedge w = v_i w_j \, dx^i \wedge dx^j \tag{4.64}$$

Finally, we define the Hodge dual of this 2-form, which is a 1-form:

$$*(v \wedge w) = \epsilon^{ij}{}_k \, v_i w_j \, dx^k \tag{4.65}$$

The components of this 1-form are

$$(v_2 w_3 - v_3 w_2, v_3 w_1 - v_1 w_3, v_1 w_2 - v_2 w_1) \tag{4.66}$$

which are precisely the components of the vector $\vec{v} \times \vec{w}$, familiar as the "cross product". It is unique to 3 dimensions that the wedge product of two vectors can again, using the Hodge dual, be expressed as a vector.

4.5 Isometry

In preceding chapters, we have defined the criteria that two spaces be identical as topological spaces (homeomorphism) and as differentiable manifolds (diffeomorphism). Now we identify the property which makes two spaces equivalent as Riemannian manifolds.

Definition: A map ϕ: $X \to Y$, where X and Y are two Riemannian manifolds, is an *isometry* if:

(i) ϕ is a homeomorphism.
(ii) ϕ is C^∞ in both directions.
(iii) The differential map $\phi_\star : T_p(X) \to T_p(Y)$ preserves the metric.

The first two requirements amount to saying that ϕ is a diffeomorphism of the two manifolds, while the last requirement says that if A and B are any two vectors in $T_p(X)$, then

$$\langle \phi_\star A, \phi_\star B \rangle_G = \langle A, B \rangle_G$$

If two Riemannian manifolds admit an isometry between them then they are said to be *isometric*, and are identical in topology, differentiable structure and metric. We consider them to be equivalent.

4.6 Integration of Differential Forms

We have seen that on an n-dimensional manifold, an n-form has a single independent component by virtue of total antisymmetry in n indices. In fact, the dual of an n-form is a 0-form or function. We can represent any n-form as:

$$\begin{aligned} \omega &= \omega_{\mu_1 \cdots \mu_n}(x)\, dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_n} \\ &= a(x)\, dx^1 \wedge \ldots \wedge dx^n \end{aligned} \tag{4.67}$$

where

$$a(x) = \epsilon^{\mu_1 \cdots \mu_n} \omega_{\mu_1 \cdots \mu_n}(x)$$

We would now like to define integration on a Riemannian manifold. For this, we assume familiarity with the integral of ordinary functions on Euclidean space, which is the same as the integral of a function of many variables over some range of its arguments. The more general integration we will now define has to be independent of the coordinate charts we use to cover the manifold, and invariant under isometries of the metric. As we will see, the natural object possessing these properties is the integral of an n-form over an n-dimensional manifold.

We will require the manifold to be orientable. This requirement can be guessed at from the fact that an integral of an ordinary function changes sign when we reverse the integration limits, or equivalently when we reverse the integration measure. The generalisation of this property will be that our integral changes sign under reversal of the orientation of the manifold.

To define the integral of an n-form on an orientable n-dimensional manifold, we recast the problem in terms of the usual integration of the function $a(x)$ that is Hodge dual to the n-form over suitably subsets of $\mathbb{R}^n$. The latter integral is just the usual Riemannian one. Combining the subsets will be done in a way that fulfils the requirements spelt out above.

Thus, we start off with a particular chart (M_α, ϕ_α) on M and define the integral of our n-form over this chart as:

$$\int_{M_\alpha} \omega = \int_{\phi_\alpha(M_\alpha) \subset \mathbb{R}^n} a(x)\, dx^1 \cdots dx^n$$

The right hand side is, as promised, the ordinary integral of a function of several variables over some range of these variables.

Note that $a(x)$ is not a scalar function on the manifold, since ω is coordinate-independent while:

$$dx^1 \wedge \ldots \wedge dx^n \to dy^1 \wedge \ldots \wedge dy^n = \left\| \frac{\partial y^\mu}{\partial x^\nu} \right\| dx^1 \wedge \cdots \wedge dx^n$$

It follows that under coordinate transformations,

$$a(x) \to a'(y) = \left\| \frac{\partial x^\mu}{\partial y^\nu} \right\| a(x)$$

or in words, $a(x)$ transforms by a multiplicative Jacobian factor. Such an object is sometimes called a *scalar density.*

This transformation of $a(x)$ precisely cancels the variation under coordinate transformations of the Cartesian integration measure $dx^1 \cdots dx^n$. Thus the integrand $a(x)\, dx^1 \cdots dx^n$ is coordinate independent. This explains why we started with an n-form, whose Hodge dual is a scalar density, instead of simply trying to integrate a scalar function.

Suppose now we take another chart (M_α, ψ_α), namely the same open set M_α but a different homeomorphism ψ_α in place of the original ϕ_α. We then see that:

$$\int_{\psi_\alpha(M_\alpha) \subset \mathbb{R}^n} a'(y)\, dy^1 \cdots dy^n = \int_{\phi_\alpha(M_\alpha) \subset \mathbf{R}^n} a(x)\, dx^1 \cdots dx^n$$

Thus, $\int_{M_\alpha} \omega$ on a *given* open set of M is invariantly defined.

It only remains to put the contributions from different open sets together to get an integral over all of M. In order not to encounter divergences in the process, we confine ourselves to n-forms with *compact support.* The support of ω is defined as:

$$\mathrm{Supp}(\omega) = \text{closure of } \{\, x \in M \mid \omega(x) \neq 0 \,\}$$

If this set is *compact*, as defined in Chapter 1, then ω is said to have compact support.

Alternatively, instead of worrying about forms with compact support we may simply confine our attention to integration over compact manifolds. Since every closed subset of a compact manifold is compact (see Chapter 1), we are then guaranteed compact support.

Finally we turn to the actual process of *patching up* the integral of ω over various open sets of M. This is achieved by defining a special collection of functions on M called a *partition of unity.* Roughly, we will decompose the constant function 1 on the manifold into a sum of functions each having support only within a given open set in the cover. This construction will enable us to patch together the integrals of a form on the various charts M_α, which we have defined above, to give an integral over all of M.

To implement this, first we need a few definitions.

Definition: An open covering $\{U_\alpha\}$ of M is said to be *locally finite* if each point $p \in M$ is contained in a finite number of U_α. A space is *paracompact* if every open cover admit a locally finite refinement. (A refinement is defined as follows: If $\{U_\alpha\}$ is an open cover, then another open cover $\{V_\beta\}$ is a *refinement* of $\{U_\alpha\}$ if every set V_β is contained in some set U_α.) Evidently, paracompactness is a weaker requirement than compactness.

Now take a paracompact manifold M (incidentally, every metric space is paracompact) and a locally finite atlas $\{U_\alpha, \phi_\alpha\}$. Pick a family of differentiable functions $e_\alpha(x)$ on M. Each e_α is chosen to be nonzero only within the corresponding open set U_α. The precise requirements are:
(i) $0 \leq e_\alpha(x) \leq 1$, for each α and all $x \in U_\alpha$.
(ii) e_α are C^∞ functions.
(iii) $e_\alpha(x) = 0$ if $x \notin U_\alpha$.
(iv) $\sum_\alpha e_\alpha(x) = 1$.
Note that the sum in (iv) contains finitely many terms for each x, precisely because the cover is locally finite.

The collection of functions e_α is called a *partition of unity subordinate to the open cover* $\{U_\alpha, \phi_\alpha\}$. Given such a partition of unity, we can finally define the integral of an n-form over the entire manifold M.

Definition: The *integral* of an n-form ω over a manifold M is:

$$\int_M \omega \equiv \sum_\alpha \int_{U_\alpha} e_\alpha \omega$$

In order for this definition to have an intrinsic meaning, it must be independent of the choice of the partition of unity and the atlas. We now prove that this is the case. In the process, the meaning of the definition should become clearer.

Given two partitions of unity, e_α subordinate to (U_α, ϕ_α) and $e'_{\alpha'}$, subordinate to $(U'_{\alpha'}, \phi'_{\alpha'})$ we have

$$\begin{aligned} \sum_\alpha \int_{U_\alpha} e_\alpha \omega &= \sum_\alpha \int_{\phi_\alpha(U_\alpha)} e_\alpha(x)\, a(x)\, dx^1 \cdots dx^n \\ &= \sum_\alpha \int_{\mathbb{R}^n} e_\alpha(x)\, a(x)\, dx^1 \cdots dx^n \end{aligned} \tag{4.68}$$

where in the second line we were able to extend the range of integration to all of $\mathbb{R}^n$ since e_α has support only in $\phi_\alpha(U_\alpha)$.

Now we multiply by $1 = \sum_{\alpha'} e'_{\alpha'}$ to get

$$\begin{aligned}
\sum_{\alpha} \int_{U_\alpha} e_\alpha \omega &= \Big(\sum_{\alpha'} e'_{\alpha'}(x') \Big) \sum_{\alpha} \int_{\mathbf{R}^n} e_\alpha(x)\, a(x)\, dx^1 \cdots dx^n \\
&= \sum_{\alpha',\alpha} \int_{\mathbf{R}^n} e_\alpha(x)\, e'_{\alpha'}(x')\, a(x)\, dx^1 \cdots dx^n \\
&= \sum_{\alpha'} \int_{\mathbf{R}^n} \Big(\sum_{\alpha} e_\alpha(x) \Big) e'_{\alpha'}(x')\, a(x')\, dx'^1 \cdots dx'^n \\
&= \sum_{\alpha'} \int_{\mathbf{R}^n} e'_{\alpha'}(x')\, a'(x')\, dx'^1 \cdots dx'^n \\
&= \sum_{\alpha'} \int_{\phi'_{\alpha'}(U'_{\alpha'})} e'_{\alpha'}(x')\, a'(x')\, dx'^1 \cdots dx'^n \\
&= \sum_{\alpha'} \int_{U'_{\alpha'}} e'_{\alpha'}\, \omega \qquad (4.69)
\end{aligned}$$

In the third step we used the coordinate independence of $a(x)\, dx^1 \cdots dx^n$ while in the last step, we used the fact that $e'_{\alpha'}$ has support in $U'_{\alpha'}$. This proves the desired result.

4.7 Stokes' Theorem

For a suitable region in $\mathbb{R}^n$, Stokes' theorem is a familiar result which relates the integral of a divergence over the region, to the integral of the corresponding function over the boundary of the region. The generalization to manifolds, which we now develop, is an important theorem and will reappear in the context of *de Rham cohomology* which we will study later on.

Let ω be an $(n-1)$-form on an n-dimensional manifold:

$$\omega = \omega_{\mu_1 \cdots \mu_{n-1}}(x)\, dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_{n-1}}$$

The Hodge dual of ω is a 1-form with components which we will denote $a^\nu(x)$. Then we can write:

$$\omega = \sum_{\nu=1}^{n} (-1)^{\nu-1} a^\nu(x)\, dx^1 \wedge \ldots \wedge \hat{dx^\nu} \wedge \ldots dx^n$$

($\hat{dx^\nu}$ means dx^ν is deleted).

Now the exterior derivative of ω is an n-form, so its dual will be a 0-form.

We have:

$$\begin{aligned} d\omega &= \sum_{\mu=1}^{n} \partial_\mu \omega_{\mu_1 \cdots \mu_{n-1}}(x)\, dx^\mu \wedge dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_n} \\ &= \sum_{\mu=1}^{n} \sum_{\nu=1}^{n} (-1)^{\nu-1} \partial_\mu a^\nu(x)\, dx^\mu \wedge dx^1 \wedge \ldots \wedge d\hat{x}^\nu \wedge \ldots \wedge dx^n \\ &= \left(\sum_\nu \partial_\nu a^\nu(x) \right) dx^1 \wedge \ldots \wedge dx^n \end{aligned} \tag{4.70}$$

(The last equation looks non-covariant: in fact, both $\partial_\nu a^\nu$ and $dx^1 \wedge \cdots \wedge dx^n$ are not coordinate independent, but the product is, since it is $d\omega$).

Thus we see that the exterior derivative of an $(n-1)$-form can be expressed in terms of the *divergence* of the components of the dual 1-form. To set up Stokes' theorem we next have to define suitable regions in a manifold.

Definition: Let M be a differentiable manifold. A subset D of M is called a *regular domain* (or simply *domain*) if for each $p \in \bar{D}$ (recall that $\bar{D}$ denotes the closure of D), either one of the following holds:
(i) There exists a local chart (U, ϕ) of M with $p \in U$, such that $\phi(\bar{D} \cap U)$ is open in $\mathbb{R}^n$, or
(ii) No such chart exists, but there exists a chart (U, ϕ) with $p \in U$ such that $\phi(\bar{D} \cap u)$ is an open set of the *half-space*

$$H = \{\, (x^1, \cdots, x^n) \in \mathbb{R}^n \mid x^n \geq 0 \,\}$$

Points $p \in D$ satisfying (i) are called *interior points* of D, while points satisfying (ii) are called *boundary points* of D. The reason for this terminology should be evident from the definition.

The set of boundary points of D is denoted ∂D and defined as

$$\partial D = \{\, p \in \bar{D} \,|\, x^n = 0 \,\}$$

It is a theorem (whose proof we skip here) that ∂D is an $(n-1)$ dimensional orientable submanifold of D.

We now have the necessary ingredients to state, and prove, Stokes' Theorem.

Theorem (Stokes): Let M be an oriented n-dimensional manifold and let D be a regular domain in M. Let ω be an $(n-1)$-form on M with compact support. Then

$$\int_D d\omega = \int_{\partial D} \omega$$

(Strictly speaking, ω is a form on M and not on ∂D. But we have an injective map, the *inclusion* i: $\partial D \to M$ which maps points in ∂D to the corresponding

points in M. Then the map i^* which we defined earlier sends forms on M to forms on ∂D. Thus, by ω on the RHS of the above, we really mean $i^*\omega$.)

Proof: Choose a locally finite cover (U_α, ϕ_α) of M and a partition of unity e_α subordinate to this cover. Then

$$\int_D d\omega = \sum_\alpha \int_D d(e_\alpha \omega)$$

On the RHS, we have a sum over forms, each of which has support in a single U_α. Thus, it is enough to verify the theorem for such forms, so we take ω to have support in one particular U_α.

Now we can prove Stokes' theorem in each of two possible situations:

Case 1: $U_\alpha \cap \partial D = \phi$ (the open set does not intersect the boundary of D). Then:

$$\int_{\partial D} \omega = 0$$

Now, since U_α does not intersect the boundary of D, it either lies entirely inside or entirely outside D. Thus $U_\alpha \subset M - \bar{D}$ or $U_\alpha \subset D$. In the former case, $\int_D d\omega = 0$ and we are done. In the latter case,

$$\int_D d\omega = \int_{U_\alpha} d\omega = \int_{\phi_\alpha(U_\alpha)} \left(\sum_\nu \partial_\nu a^\nu(x) \right) dx^1 \cdots dx^n$$

Since ω has compact support, we can take the integration region to be a *cube* in $\mathbb{R}^n$, call it C, of side 2λ, such that $a_\mu = 0$ on the border of C (any $x^\mu = \lambda$). Then

$$\begin{aligned} \int_D d\omega &= \int_C \left(\sum_\nu \partial_\nu a^\nu(x) \right) dx^1 \cdots dx^n \\ &= \sum_\nu \int_C a^\nu(x^1 \cdots x^n)\, dx^1 \cdots d\hat{x}^\nu \cdots dx^n \Big|_{x_\nu = -\lambda}^{x_\nu = \lambda} = 0 \end{aligned} \tag{4.71}$$

Thus we have shown that both sides of Stokes' theorem vanish in this case, hence the theorem is true.

Case 2: $U_\alpha \cap \partial D \neq \phi$. Then, since ∂D is defined by $x^n = 0$, we have

$$\int_{\partial D} \omega = (-1)^{n-1} \int a_n(x^1, \cdots, x^{n-1}, 0)\, dx^1 \cdots dx^{n-1}$$

and

$$\int_D d\omega = \int_C \left(\sum_\nu \partial_\nu a^\nu \right) dx^1 \cdots dx^n$$

where now the cube extends from $-\lambda$ to λ in all directions *except* the n^{th} one, where it extends from 0 to λ. Thus, for any fixed $j \neq n$ we have:

$$\int_C (\partial_\nu a^\nu) dx^1 \cdots dx^n = 0$$

(there is no sum over ν in the above equation), while:

$$\int_C (\partial_n a_n)\, dx^1 \cdots dx^n = (-1)^{n-1} \int a_n(x^1, \cdots, x^{n-1}, 0)\, dx^1 \cdots dx^{n-1}$$

Thus we have shown that both sides of Stokes' theorem are equal to the same expression. This finally proves the theorem.

4.8 The Laplacian on Forms

For a p-form $\omega = \omega_{\mu_1 \cdots \mu_p}(x)\, dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_p}$, we have already defined the Hodge dual:

$$^*\omega = \frac{\sqrt{g}}{(n-p)!} \epsilon^{\mu_1 \cdots \mu_p}{}_{\mu_{p+1} \cdots \mu_n}\, \omega_{\mu_1 \cdots \mu_p}(x)\, dx^{\mu_{p+1}} \wedge \ldots \wedge dx^{\mu_n}$$

The Hodge dual enables us to define an *inner product* on forms as follows:

$$\langle \alpha, \beta \rangle = \int_M \alpha \wedge {}^*\beta$$

Since α and β are p-forms, $\alpha \wedge {}^*\beta$ is an n-form, hence as we have just seen, it makes sense to integrate it over the manifold M (we assume that M has no boundary.) The inner product thus takes two p-forms, for any p, and gives a number.

In components,

$$\alpha \wedge {}^*\beta = \sqrt{g}\, g^{\mu_1 \nu_1} \cdots g^{\mu_p \nu_p}\, \alpha_{\mu_1 \cdots \mu_p}\, \beta_{\nu_1 \cdots \nu_p}\, dx^1 \wedge \cdots \wedge dx^n.$$

so the inner product of the forms is the integral over the inner product of the components, with all indices contracted via the metric, and with a weight factor $\sqrt{g}$ which ensures coordinate-independence of the result.

From the above equation it is clear that the inner product is symmetric:

$$\langle \alpha, \beta \rangle = \langle \beta, \alpha \rangle$$

and moreover, the inner product of a form with itself is greater than or equal to zero:

$$\langle \alpha, \alpha \rangle \geq 0$$

Since the integral of a strictly positive quantity can vanish only if the quantity itself vanishes, this can be zero only if the the square of the form components is zero at each point, which in turn means the form itself is identically zero.

Given the exterior derivative, the inner product allows us to define its *adjoint*, denoted δ. This is defined by

$$\langle \alpha, d\beta \rangle = \langle \delta\alpha, \beta \rangle.$$

Going to components, one finds

$$\int_M \alpha \wedge {}^*d\beta = (-1)^{np} \int_M ({}^*d{}^*\alpha) \wedge {}^*\beta$$

where the sign depends on the product of n (the dimension of the manifold) and p (the dimension of the form).

Thus, the adjoint operator δ can be thought of as $\pm {}^*d^*$. The action of δ *lowers* the dimension of a form by one unit, the opposite of what d does. If ω is a p-form, then ${}^*\omega$ is an $(n-p)$ form, $d^*\omega$ is an $(n-p+1)$ form and ${}^*d^*\omega$ is an $(n-(n-p+1)) = (p-1)$ form.

Example: In the particular case of three space dimensions ($n = 3$), we find that the d-operator on 1-forms is the familiar curl:

$$d: \ \omega = \omega_i \, dx^i \rightarrow d\omega = \frac{1}{2}(\partial_i \omega_j - \partial_j \omega_i) \, dx^i \wedge dx^j$$

We can now check that δ on 1-forms is the *divergence*:

$$\delta: \ \omega = \omega_i \, dx^i \rightarrow \frac{1}{\sqrt{g}} \partial_i \left(\sqrt{g} \, \omega^i \right) \tag{4.72}$$

Exercise: Derive the above result and show that the RHS is a genuine scalar under coordinate transformations.

Clearly $\delta\omega$ defines the generalization of the divergence $\partial_i \, \omega^i$ to an arbitrary Riemannian manifold in 3 dimensions.

Just as $d^2 = 0$, we also have $\delta^2 = 0$:

$$\begin{aligned} \delta\delta\omega &= \pm({}^*d^*)({}^*d^*)\,\omega \\ &= \pm({}^*d^2{}^*)\,\omega \\ &= 0 \end{aligned} \tag{4.73}$$

A very important second-order differential operator in familiar, flat space physics is the *Laplacian* $\Delta = \partial_i \, \partial_i$. The generalization to a manifold is defined using the d and δ operators.

Definition: The Laplacian Δ is defined by

$$\Delta \equiv (d+\delta)^2 = d\delta + \delta d$$

The equality of the two expressions is of course due to $d^2 = \delta^2 = 0$.

Clearly the Laplacian (unlike d and δ) maps p-forms back to p-forms:

$$\Delta: \ \wedge^p(M) \rightarrow \wedge^p(M)$$

Moreover it is self-adjoint in the norm on p-forms. Thus, for example, it makes sense to talk of eigenvectors and eigenvalues of the Laplacian on a manifold. The study of these is called *harmonic analysis*.

In the special case of *Euclidean space* $\mathbb{R}^n$ and in Cartesian coordinates, the Laplacian on any p-form gives:

$$\Delta\omega = \left(\partial_j \partial_j \, \omega_{i_1 \cdots i_p}(x) \right) dx^{i_1} \wedge \cdots \wedge dx^{i_p}$$

Thus in this case it is indeed the usual Laplacian $\partial_i \partial_i$ on the individual components. The Laplacian is a *positive* operator, which means its matrix elements between any pair of forms is greater than or equal to zero:

$$\begin{aligned} \langle \omega, \Delta\omega \rangle &= \langle \omega, d\delta\omega + \delta d\omega \rangle \\ &= \langle \delta\omega, \delta\omega \rangle + \langle d\omega, d\omega \rangle \\ &\geq 0. \end{aligned} \tag{4.74}$$

It also follows from this that $\Delta\omega = 0$ if and only if $\delta\omega = 0$ *and* $d\omega = 0$

At this point we introduce some terminology and then state an important theorem.

Definition: A form ω is said to be *closed* if $d\omega = 0$, and *exact* if $\omega = df$ for some other form f. Similarly, ω is said to be *co-closed* if $\delta\omega = 0$ and *co-exact* if $\omega = \delta f$ for some other form f. Finally, a form satisfying $\Delta\omega = 0$ is said to be *harmonic*. Clearly a form is harmonic if and only if it is both closed and co-closed.

Hodge decomposition theorem: If M is a compact manifold without boundary, then any p-form ω can be uniquely decomposed as the sum of exact, co-exact and harmonic forms:

$$\omega = d\alpha + \delta\beta + \gamma$$

for some forms α, β, γ, where $\Delta\gamma = 0$. Clearly if ω is a p-form, then α, β, γ are $(p-1), (p+1)$ and p forms respectively.

We omit the proof of this theorem here as it is somewhat involved. Nevertheless the result will be a useful tool in studying de Rham cohomology in the following chapter.

Chapter 5

Homology and Cohomology

5.1 Simplicial Homology

We have seen in Chapter 2 that the topological properties any topological space can be understood to some extent via *homotopy*, the study of loops in the space. An alternative approach to studying topological properties, for a differentiable manifold, arises through the study of objects called *simplices*. This can be used to characterize topological properties of manifolds in terms of *simplicial homology*. A closely related methodology, though from a completely different starting point, is to characterise topology through the study of differential forms. This goes by the name of *de Rham cohomology*. The two approaches are in a certain sense dual to each other, as we will explain.

We start by developing the theory of simplices and simplicial complexes. Intuitively, the idea is to define nice subsets of Euclidean space which look like *polyhedra*. These can be used to cover a manifold by a process called *triangulation*. Topological properties of the manifold can then be expressed in terms of the pieces which triangulate it.

To develop some intuition about this process, a simple example is provided by taking the 2-sphere S^2 and drawing triangles all over it to completely cover it. Each of the triangles has one edge in common with some other triangle (Fig. 5.1)

The triangles in the figure are triangles only in the sense that they are bounded by three lines. These are not "straight lines" in any sense, and the answers we extract will not depend on any local details of the lines or triangles. The important thing, as we will see in what follows, is that a generalised version of triangulation provides a powerful tool for the formulation and solution of problems regarding the topological properties of a manifold in any number of dimensions.

For the general case, we first define an object called a simplex. This is defined in Euclidean space $\mathbb{R}^n$ with no reference to the manifold of interest. If $x_1, x_2, \cdots, x_{M+1}$ are distinct points in $\mathbb{R}^n$, they are said to be *linearly indepen-*

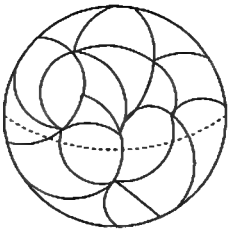


Figure 5.1: A triangulation of S^2.

dent if the M vectors $x_2 - x_1, x_3 - x_1, \cdots, x_{M+1} - x_1$ are linearly independent vectors. A simplex will be obtained by "filling in" the region in $\mathbb{R}^n$ defined by a set of such points.

Definition: An *M-simplex* σ^M is the set of points

$$\sigma^M = \left\{ x \in \mathbb{R}^n \mid x = \sum_{i=1}^{M+1} \lambda_i x_i, \quad \sum_{i=1}^{M+1} \lambda_i = 1, \quad \lambda_i \geq 0 \right\} \tag{5.1}$$

where $x_1, \cdots, x_{M+1}$ are linearly independent as defined above.

Here are some examples of simplices:
(i) A 1-simplex in $\mathbb{R}^3$ is the set

$$x = \lambda_1 x_1 + \lambda_2 x_2 \tag{5.2}$$

with $\lambda_1 + \lambda_2 = 1$. Thus, it is the collection of points:

$$x = \lambda_1 x_1 + (1 - \lambda_1) x_2 \tag{5.3}$$

as λ_1 varies from 0 to 1. Clearly, this is the straight line joining x_1 to x_2.
(ii) A 2-simplex in $\mathbb{R}^3$ is the set

$$x = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 \tag{5.4}$$

with $\lambda_1 + \lambda_2 + \lambda_3 = 1$. Thus,

$$x = \lambda_1 x_1 + \lambda_2 x_2 + (1 - \lambda_1 - \lambda_2) x_3 \tag{5.5}$$

This is just the interior of the *triangle*, including the boundary. Linear independence of the points x_1, x_2, x_3 ensures that they are not collinear.

In general, the points $\{x_1, \cdots, x_{M+1}\}$ are called the *vertices* of the M-simplex σ^M, which itself is denoted

$$\sigma^M \equiv [x_1, \cdots, x_{M+1}] \tag{5.6}$$

Note that an M-simplex is not just the set of vertices but the whole region "enclosed" by them. What we have defined are more accurately described as

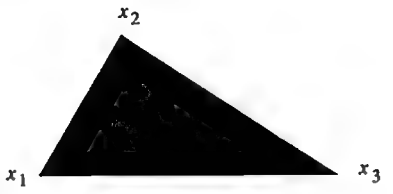


Figure 5.2: The shaded region (including the boundary) is an example of a 2-simplex.

"closed simplices". It is evident that each M-simplex is homeomorphic to $[0,1]\otimes [0,1]\otimes\cdots\otimes[0,1]$ (M times), or in other words, to a closed subset of $\mathbb{R}^M$.

The set of numbers λ_i which label points in the simplex are called *barycentric coordinates*. To understand this terminology, observe that if masses λ_i, $i = 1,\cdots,M+1$ are placed at the points x_i, $(i = 1,\cdots,M+1)$ then the centre of mass is located at $x = \sum_{i=1}^{M+1}\lambda_i x_i$, if $\sum_i \lambda_i = 1$.

For each M-simplex, we can define the *faces* as the collection of $(M-1)$-simplices which lie "opposite" the vertices. For example, for a triangle, we can associate to each vertex the line that lies opposite it. Thus we have:

Definition: The *jth face* of an M-simplex σ^M is the $(M-1)$-simplex $\sum_{i=1,\ i\neq j}^{M+1}\lambda_i x_i$ with $\sum_{i=1,\ i\neq j}^{M+1}\lambda_i = 1$. This amounts to considering the subset of the M-simplex obtained by setting $\lambda_j = 0$. Since an M-simplex has $M+1$ vertices, it evidently also has $M+1$ faces.

Definition: An *open simplex* is the interior of any closed simplex. We denote this by (σ^M). Given independent points $x_1, x_2, \cdots, x_{M+1}$, an open simplex is

$$\sigma^M = \{\ x\in\mathbb{R}^n \mid x = \sum_{i=1}^{M+1}\lambda_i x_i,\quad \sum_{i=1}^{M+1}\lambda_i = 1,\quad \lambda_i > 0\ \} \tag{5.7}$$

We have simply modified the definition of closed simplex by restricting to $\lambda_i > 0$ instead of $\lambda_i \geq 0$. Clearly an open simplex is an open set in the corresponding Euclidean space.

Finally, we define a simplicial complex as a collection of simplices with specific properties.

Definition: A *simplical complex* K is a finite collection of (closed) simplices in some $\mathbb{R}^n$ satisfying:
(i) If $\sigma^p \in K$, then all faces of σ^p belong to K.
(ii) If $\sigma^p, \sigma^q \in K$, then either $(\sigma^p)\cap(\sigma^q) = \phi$, or $\sigma^p = \sigma^q$.

(Recall that by (σ^p) we mean the open simplex corresponding to the closed simplex σ^p.) The *dimension* of a simplicial complex K is defined to be the dimension of $\sigma^p \in K$ for the largest p.

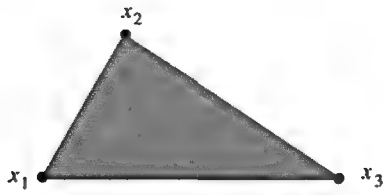


Figure 5.3: A two-dimensional simplical complex which is the union of three 0-simplices, three 1-simplices and one 2-simplex.

Basically the first part of the definition says that if a particular simplex belongs to a simplicial complex, then that complex must necessarily contain all the lower dimensional simplices in the original one. The second part says that any two distinct simplices in the complex can only touch along components of their faces, but cannot actually overlap each other.

An example of a 2-dimensional simplicial complex can be defined by its collection of simplices as follows:

$$K = \{\ [x_1, x_2, x_3], [x_1, x_2], [x_2, x_3], [x_1, x_3], [x_1], [x_2], [x_3]\ \} \tag{5.8}$$

This complex is illustrated in Fig. 5.3.

Some more possible two-dimensional simplicial complexes are illustrated in Fig. 5.4, while Fig. 5.5 shows some objects that are *not* simplicial complexes. For example, in the first two diagrams of Fig. 5.5 a 0-simplex is missing. (In these figures, heavy dots indicate 0-simplices.)

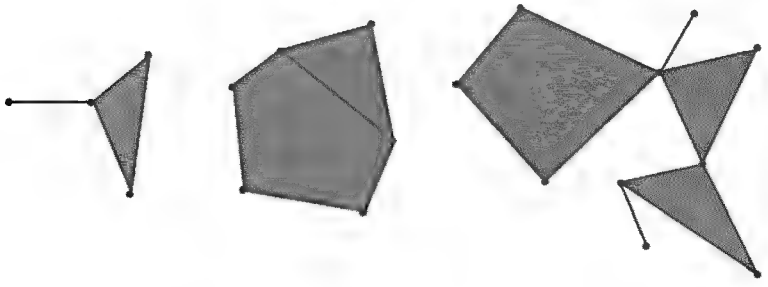


Figure 5.4: Examples of simplicial complexes (check that the axioms hold!).

We still need a few more definitions before we can use simplicial complexes

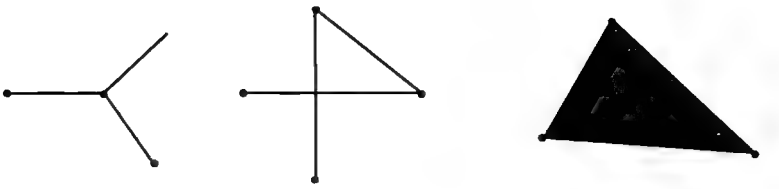


Figure 5.5: These are *not* simplicial complexes (0-simplices and 1-simplices are missing in some of these diagrams).

to study the topology of differentiable manifolds.

Definition: The union of all members of a simplicial complex K with the Euclidean subspace topology is called the *polyhedron* associated with K. (The polyhedron will also be denoted K, whenever there is no chance of confusion.)

Definition: A *smooth triangulation* of a differentiable manifold M is a homeomorphism $\phi : K \to M$ for some polyhedron K. (ϕ must satisfy a technical property that we will not go into here.)

So far, this has been a rather abstract discussion with no reference to manifolds. The key result that relates all this to the topic of our interest is that *every compact C^∞ manifold can be smoothly triangulated.* This is the basis of the study of manifolds through simplicial complexes. We will not be able to provide a proof of this result here.

We will need to refine the definition of a simplex to take into account the concept of *orientation*:

Definition: An *oriented* p-simplex ($p \geq 1$) is a p-simplex along with an ordering for its vertices. The equivalence class of even permutations of the chosen ordering defines a positively oriented simplex $+\sigma^p$, while odd permutations give the negatively oriented simplex, denoted $-\sigma^p$. (Oriented simplices should be denoted by a new symbol, such as $\langle\sigma^p\rangle$, but we avoid it to save notation and because henceforth we will always work with oriented simplices).

For example, a 2-simplex $\sigma^2 = [V_0, V_1, V_2]$ is associated to the positively oriented simplex

$$+\sigma^2 = [V_0, V_1, V_2] = [V_1, V_2, V_0] = [V_2, V_0, V_1] \tag{5.9}$$

and the negatively oriented simplex

$$-\sigma^2 = [V_0, V_2, V_1] = [V_2, V_1, V_0] = [V_1, V_0, V_2] \tag{5.10}$$

Associated to every oriented p-simplex, we would like to define a set of oriented $(p-1)$ simplices called the *boundary* of σ^p. As a set of simplices, the boundary is just the collection of faces of σ^p. But we need to assign an

orientation to each face. This can be motivated as follows: if $[V_0, V_1]$ is an ordered 1-simplex, then the boundary (denoted by ∂) must satisfy:

$$\partial[V_0, V_1] = -\partial[V_1, V_0] \tag{5.11}$$

At the same time, the boundary will naturally contain just the two end points of this one-dimensional simplex, namely $[V_0]$ and $[V_1]$.

Clearly, the only possibility satisfying these conditions is:

$$\partial[V_0, V_1] = [V_1] - [V_0] \tag{5.12}$$

(or the opposite signs on the right, but that is an overall convention). The *sum* of two oriented 0-simplices on the right hand side is a formal one, and we will give it a precise meaning below.

Generalising this idea to one higher dimension, we see that the boundary of an oriented 2-simplex is

$$\partial[V_0, V_1, V_2] = [V_1, V_2] - [V_0, V_2] + [V_0, V_1] \tag{5.13}$$

The rule is evidently to eliminate one vertex at a time and to choose a sign corresponding to the location of the deleted vertex. Thus for a general p-simplex, we have:

Definition: The *boundary* of an oriented p-simplex is

$$\partial[V_0, V_1, \cdots, V_p] = \sum_{j=0}^{p} (-1)^j [V_0, \cdots, \hat{V}_j, \cdots, V_p] \tag{5.14}$$

where $\hat{V}_j$ means that the vertex V_j is omitted in that term.

At this stage we need to define what the formal sums used above really mean. Actually, all we will do is to give a name to such sums of simplices, and enlarge our notion of simplices to include such formal sums. This will give a useful group structure to the study of simplicial complexes.

Definition: Let K be an n-dimensional simplicial complex, containing l_p p-simplices. A *p-chain* of K is a formal sum of oriented simplices with integer coefficients. Thus, if σ_i^p $(i = 1, \cdots, l_p)$ are oriented p-simplices, then

$$C_p = \sum_{i=1}^{l_p} n_i \sigma_i^p, \qquad n_i \in \mathbb{Z} \tag{5.15}$$

is a p-chain. So in fact the boundary of a p-simplex, as defined above, is a $(p-1)$-chain. In terms of chains, a simplex can be thought of as the special case of a chain where there is only one term in the sum.

The boundary operator on p-chains is easily defined in terms of the boundary operator on each simplex in the chain:

$$\partial C_p = \partial \left(\sum_{i=1}^{l_p} n_i \sigma_i^p \right) \equiv \sum_{i=1}^{l_p} n_i (\partial \sigma_i^p) \tag{5.16}$$

The collection of p-chains forms an *abelian group*. The group axioms are established as follows:

(i) $C_p = \sum n_i \sigma_i^p$, $D_p = \sum m_i \sigma_i^p \Rightarrow C_p + D_p = \sum (n_i + m_i)\sigma_i^p$.

(ii) The identity chain 0 is obtained by picking all $n_i = 0$.

(iii) If $C_p = \sum n_i \sigma_i^p$ is a chain then $-C_p \equiv \sum (-n_i)\sigma_i^p$ is another chain, with $C_p + (-C_p) = 0$.

(iv) Associativity obviously holds.

This is called the *free Abelian group* generated by the p-simplices of K, denoted $C_p(K, \mathbb{Z})$ or simply $C_p(K)$. (The $\mathbb{Z}$ denotes that we only consider integer multiples, which is part of the meaning of free Abelian group).

Summarising the discussion above, we have defined a map ∂, called the boundary map, on p-chains. (To be precise, we will label the map ∂_p when it acts on p-chains, though this may seem a bit pedantic. The reader may have realised that even for the exterior derivative operator d, which we discussed earlier, one could assign a label p when it acts on p-forms. It is necessary to stress this point only when confusion might otherwise arise). This map:

$$\partial_p \; : \; C_p(K) \rightarrow C_{p-1}(K)$$

has the following properties:

(i) $\partial_p \left(\sum n_i \sigma_i^p \right) = \sum n_i (\partial_p \sigma_i^p)$

(ii) $\partial_p \sigma^p = \partial_p [V_0, V_1, \cdots, V_p] \equiv \sum_{j=0}^{p} (-1)^j [V_0, \cdots, \hat{V}_j, \cdots, V_p]$

(iii) $\partial_p [V_0] = 0$, since a 0-simplex cannot have a boundary.

(iv) ∂_p is a homomorphism of the Abelian groups $C_p(K) \rightarrow C_{p-1}(K)$.

For the last point, one can check for example that $\partial_p(-C_p) = -\partial_p(C_p)$, $\partial_p(C_p + D_p) = \partial_p C_p + \partial_p D_p$, and so on. However, it is important to note that this homomorphism is *not* an isomorphism, because in general it is many-to-one. For example, two different chains may have a vanishing boundary, in which case they are both mapped to $0 \in C_{p-1}$. More generally, two or more chains can have the same boundary.

As an example, any 2-chain of the form $[V_0, V_1] + [V_1, V_2] + [V_2, V_0]$ satisfies

$$\begin{aligned} \partial_2 \Big([V_0, V_1] + [V_1, V_2] + [V_2, V_0] \Big) &= \Big([V_1] - [V_0] \Big) + \Big([V_2] - [V_1] \Big) + \Big([V_0] - [V_2] \Big) \\ &= 0 \end{aligned} \tag{5.17}$$

Now we will prove an important theorem about the boundary operator ∂: its square is zero, just as for the exterior derivative.

Theorem: $\partial^2 = 0$ (this really means $\partial_{p-1} \partial_p = 0$ in our pedantic notation).

Proof: The proof is straightforward if slightly tedious to write down. It basically relies on a kind of antisymmetry. We have:

$$\begin{aligned}
\partial_{p-1}(\partial_p \sigma^p) &= \partial_{p-1}\left(\sum_{j=0}^{p}(-1)^j[V_0,\cdots,\hat{V}_j,\cdots,V_p]\right) \\
&= \sum_{j=0}^{p}(-1)^j\partial_{p-1}[V_0,\cdots,\hat{V}_j,\cdots,V_p] \\
&= \sum_{j=0}^{p}(-1)^j\left(\sum_{i=0}^{j-1}(-1)^i[V_0,\cdots,\hat{V}_i,\cdots,\hat{V}_j,\cdots,V_p]\right. \\
&\qquad \left. + \sum_{i=j+1}^{p}(-1)^{i-1}[V_0,\cdots,\hat{V}_j,\cdots,\hat{V}_i,\cdots,V_p]\right) \\
&= \sum_{i<j}(-1)^{i+j}[V_0,\cdots,\hat{V}_i,\cdots,\hat{V}_j,\cdots,V_p] \\
&\qquad + \sum_{i>j}(-1)^{i+j-1}[V_0,\cdots,\hat{V}_j,\cdots,\hat{V}_i,\cdots,V_p] \\
&= \sum_{i<j}\left((-1)^{i+j}+(-1)^{i+j-1}\right)[V_0,\cdots,\hat{V}_i,\cdots,\hat{V}_j,\cdots,V_p] \\
&= 0
\end{aligned} \tag{5.18}$$

The last equality follows since $(-1)^{i+j}+(-1)^{i+j-1}=0$.

Using these properties of the boundary operator, we can identify two useful kinds of chains.

(i) If $\partial C_p = 0$ then the chain C_p is called a *cycle*.

(ii) If $C_p = \partial C_{p+1}$ for some chain C_{p+1}, then C_p is called a *boundary*.

Clearly, if $C_p = \partial C_{p+1}$ then $\partial C_p = 0$. Thus every boundary is a cycle. However, the reverse is not true in general, as we will see. Indeed, if the boundary of a chain is zero, that does not necessarily imply that this chain is a boundary of some other one. This property turns out to capture a lot of information about the manifold which will be triangulated using simplicial complexes.

We have already noted that the p-chains of K form an Abelian group $C_p(K)$. It is easy to see that cycles and boundaries form Abelian subgroups of this group. Define

$$\begin{aligned}
Z_p(K) &\equiv \{\ C_p \in C_p(K) \mid \partial C_p = 0\ \} \\
B_p(K) &\equiv \{\ C_p \in C_p(K) \mid C_p = \partial C_{p+1}\ \}
\end{aligned} \tag{5.19}$$

Thus $Z_p(K)$ and $B_p(K)$ are, respectively, the groups formed by the cycles and the boundaries of K, and clearly, $B_p(K) \subset Z_p(K) \subset C_p(K)$ (not just as sets but also as groups).

Now we are in a position to define the concept of a homology group. This is again an Abelian group, in which the elements are the cycles up to possible addition of boundaries.

Definition: The p-th *homology group* $H_p(K)$ of the simplicial complex K with integer coefficients, is the quotient of the group of cycles by the group of boundaries:

$$H_p(K) = Z_p(K)/B_p(K)$$

Thus, the elements of $H_p(K)$ are the *equivalence classes* of cycles which differ by boundaries. It is obvious that H is itself an Abelian group. Evidently, for an n-dimensional simplicial complex K, we have $H_p(K) = 0$ for $p > n$ since there are no chains of dimension larger than that of the complex, by definition. Thus there is a finite number of homology groups for each manifold, one for each integer p between 0 and n.

Let us make this definition still more explicit. Suppose $h \in H_p(K)$. Then h is an equivalence class of p-cycles: $h = [C_p^1, C_p^2, C_p^3, \cdots]$. Any pair of cycles C_p^i, C_p^j in this class have the property that their difference is a boundary: $C_p^i - C_p^j = \partial C_{p+1}$ for some chain C_{p+1}. A different element $h' \in H_p(K)$ will represent a *distinct* equivalent class: $h' = [D_p^1, D_p^2, D_p^3, \cdots]$. If h' is not the same class as h, it means that there is *no* C_{p+1} such that $C_p^i - D_p^j = \partial C_{p+1}$.

We can equivalently define $H_p(K)$ in terms of an "exact sequence".

Definition: If $f : G \to H$ is a homomorphism from the group G to the group H, then:

(i) The *kernel* of f, denoted $ker(f)$, is the subgroup of elements of G which get mapped to the identity in H:

$$\ker(f) \equiv \{ x \in G \mid f(x) = 0 \in H \} \tag{5.20}$$

(ii) The *image* of f, denoted $im(f)$, is the subgroup of elements of H which come from some element of G under the map:

$$\mathrm{im}(f) \equiv \{ y \in H \mid y = f(x), x \in G \} \tag{5.21}$$

Now clearly, the group $Z_p(K)$ of cycles is the same as $ker(\partial_p)$, while the group $B_p(K)$ of boundaries is the same as $im(\partial_{p+1})$. In fact, the boundary operator ∂ provides a sequence of homomorphisms

$$\cdots \overset{\partial_{p-2}}{\leftarrow} C_{p-2}(K) \overset{\partial_{p-1}}{\leftarrow} C_{p-1}(K) \overset{\partial_p}{\leftarrow} C_p(K) \overset{\partial_{p+1}}{\leftarrow} \cdots \tag{5.22}$$

This is called an *exact sequence*, which means that the kernel of the map at each stage is contained in the image of the subsequent map. This is obviously a consequence of $\partial_{p-1}\partial_p = 0$ for each p. In this language, the homology groups $H_p(K)$ defined earlier are just the quotients of the kernel by the image:

$$H_p(K) = \frac{ker\,(\partial_p)}{im\,(\partial_{p+1})} \tag{5.23}$$

So far, the homology group appears to be a property associated to a given simplicial complex (equivalently, polyhedron) K, which specifies a triangulation of the manifold. But the following result tells us that it is independent of the triangulation.

Theorem: If two polyhedra K, L are homeomorphic as topological spaces, then $H_p(K) = H_p(L)$. (We omit the proof.)

Since every compact manifold M can be smoothly triangulated by some polyhedron K, we can *define* $H_p(M)$, the homology group of the manifold, to be the homology group $H_p(K)$ for the polyhedron triangulating it. This is an important way to characterize the topology of a manifold.

Example: Let us find H_0 and H_1 for the circle, S^1. The circle can be triangulated as in Fig. 5.6. Thus,

$$K = \{\ [V_0, V_1],\ [V_1, V_2],\ [V_2, V_0],\ [V_0],\ [V_1],\ [V_2]\ \} \tag{5.24}$$

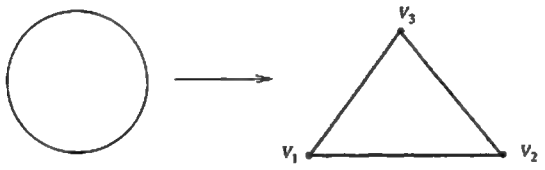


Figure 5.6: Triangulation of S^1.

Now let us look for the 1-cycles which constitute $Z_1(K)$. The most general 1-chain is:

$$C_1 = a[V_0, V_1] + b[V_1, V_2] + c[V_2, V_0] \tag{5.25}$$

where a, b, c are integers. This chain will be a cycle if $\partial C_1 = 0$. This leads to:

$$a[V_1] - a[V_0] + b[V_2] - b[V_1] + c[V_0] - c[V_2] = 0 \tag{5.26}$$

It follows that $a = b = c$. So $Z_1(K)$ is the group of elements:

$$Z_1(K) = \{\ a([V_0, V_1] + [V_1, V_2] + [V_2, V_0])\ |\ a \in \mathbb{Z}\ \} \tag{5.27}$$

Next, we look for 1-boundaries. By definition, these must bound 2-chains, but there are no 2-chains in our complex, since it is 1-dimensional. It follows that $B_1(K) = 0$. Thus,

$$\begin{aligned} H_1(K) &\equiv Z_1(K)/B_1(K) = Z_1(K) \\ &= \{\ a([V_0, V_1] + [V_1, V_2] + [V_2, V_0])\ |\ a \in \mathbb{Z}\ \} \end{aligned} \tag{5.28}$$

which is isomorphic to the group $\mathbb{Z}$ of integers. Thus we can write:

$$H_1(S^1) = \mathbb{Z} \tag{5.29}$$

For $H_0(K)$, we start by finding $Z_0(K)$. This is defined by

$$\partial(a[V_0] + b[V_1] + c[V_2]) = 0 \tag{5.30}$$

But this is always true since $\partial[V_i] = 0$ by definition. (A 0-chain cannot have a boundary.) So

$$Z_0(K) = \{\ a[V_0] + b[V_1] + c[V_2] \mid a, b, c \in \mathbb{Z}\ \} \tag{5.31}$$

which is isomorphic to $\mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$.

Next we find the 0-boundaries:

$$B_0(K) = \{\ C_0 \in C_0(K) \mid C_0 = \partial C_1\ \} \tag{5.32}$$

Take a general 1-chain, as above, then its boundary is

$$\partial\,(a[V_0, V_1] + b[V_1, V_2] + c[V_2, V_0]) = (a-b)[V_1] + (b-c)[V_2] + (c-a)[V_0] \tag{5.33}$$

Thus, an element of $B_0(K)$ is specified by three integers $(a-b)$, $(b-c)$, $(c-a)$. But these are not independent (they sum to zero!). However any two of them are independent, so

$$B_0(K) = \mathbb{Z} \oplus \mathbb{Z} \tag{5.34}$$

The quotient is obtained as follows: a general element of $Z_0(K)$ is:

$$\begin{aligned} a[V_0] + b[V_1] + c[V_2] &= (a+b+c)[V_0] + \Big((-b-c)[V_0] + b[V_1] + c[V_2]\Big) \\ &= (a+b+c)[V_0] \qquad (\text{mod } B_0(K)) \end{aligned} \tag{5.35}$$

since the second term, $\Big((-b-c)[V_0] + b[V_1] + c[V_2]\Big)$, is an element of $B_0(K)$ as we have shown above. So $H_0(K)$ is labelled by a single integer $(a+b+c)$, and hence

$$H_0(S^1) = \mathbb{Z} \tag{5.36}$$

This completes the calculation of the simplicial homology groups for S^1.

The answer that we found for $H_0(S^1)$ is actually a special case of a general result:

Theorem: If K is a connected polyhedron then $H_0(K) = \mathbb{Z}$. (The proof is straightforward.)

The homology groups defined above were built up using formal sums of simplices with integer coefficients, so they should actually be denoted $H_p(M, \mathbb{Z})$. Instead of integer coefficients, we could have considered formal sums of simplices with *real* coefficients, which define homology groups which are denoted $H_p(M, \mathbb{R})$. These groups are related, but not in a one-to-one fashion, to $H_p(M, \mathbb{Z})$.

For this, first note that for any manifold, $H_p(M, \mathbb{Z})$ must be a direct sum of Abelian groups of integers, which means it has the general form

$$H_p(M, \mathbb{Z}) = \mathbb{Z} \oplus \ldots \mathbb{Z} \oplus \mathbb{Z}_{p_1} \oplus \mathbb{Z}_{p_2} \oplus \ldots \tag{5.37}$$

where $\mathbb{Z}$ is the group of integers under addition, while $\mathbb{Z}_p$ is the finite group of integers under addition modulo p, known as the *cyclic group of order p*.

The finite groups Z_{p_i} appearing in $H_p(M, \mathbb{Z})$ are known as *torsion subgroups* of the homology group. If we consider now the real, rather than integer, cohomology, the result turns out to be:

$$H_p(M, \mathbb{R}) = \mathbb{R} \oplus \ldots \oplus \mathbb{R} \tag{5.38}$$

where the number of copies of $\mathbb{R}$ in $H_p(M, \mathbb{R})$ is the same as the number of copies of $\mathbb{Z}$ in $H_p(M, \mathbb{Z})$. In other words, in the *real* cohomology, the torsion subgroups are ignored, and only the $\mathbb{Z}$ factors in the integer homology contribute.

Definition: The number of $\mathbb{Z}$ factors in $H_p(M, \mathbb{Z})$, or equivalently the dimension of $H_p(M, \mathbb{R})$, is called the p-th *Betti number* of the manifold M, which we denote β_p.

The Betti numbers are topological invariants of a manifold, which means that any two manifolds that are homeomorphic to each other will necessarily have the same Betti numbers. However the converse is not true: having the same Betti numbers does not tell us that the two manifolds are homeomorphic.

The sum of all the Betti numbers with alternating signs is called the *Euler characteristic* $\chi(M)$ of the manifold:

$$\chi(M) = \sum_{p=0}^{\dim M} (-1)^p \beta_p \tag{5.39}$$

The Euler characteristic, like the Betti numbers in terms of which we defined it, captures something of the topology of a manifold. A very interesting theorem tells us that the same topological information can be captured merely by counting the simplices in a triangulation of the manifold.

Theorem: For an n-dimensional simplical complex K, let α_p denote the number of p-simplices of K. Then

$$\chi(K) = \sum_{p=0}^{n} (-1)^p \alpha_p \tag{5.40}$$

Thus one can deduce the Euler characteristic of a manifold merely by triangulating it, without having to actually compute the homology groups! It is *not* the case in general that $\alpha_p = \beta_p$ for each p, so that counting simplices does not determine the individual Betti numbers. Indeed, the number of simplices in a

triangulation is not a topological invariant of the manifold, since there are many different ways to triangulate a manifold with various numbers of simplices. The theorem above tells us that however we triangulate the manifold, the alternating sum of the number of simplices will come out the same and will be a topological invariant equal to the Euler characteristic.

Example: Consider the 2-sphere, S^2. This is homeomorphic to a tetrahedron, for which we easily find

$$\begin{aligned} \alpha_0 &= 4 \\ \alpha_1 &= 6 \\ \alpha_2 &= 4 \end{aligned} \tag{5.41}$$

from which it follows that

$$\chi(S^2) = 4 - 6 + 4 = 2 \tag{5.42}$$

This example is a special case of a famous relation, valid for the surface of any polyhedron in 3 dimensional space. Any polyhedron P is homeomorphic to S^2, so $\chi(P) = 2$. Now let V be the number of vertices, E the number of edges and F the number of faces of the polyhedron. In principle these numbers cannot yet be related to α_i, $i = 0, 1, 2$ since the α's refer to the simplices in a *triangulation*, while a polyhedron may have faces which are not triangles. Nevertheless, any polygonal face of a tetrahedron can be subdivided into triangles, and it is easy to check that the number $V - E + F$ for the polyhedron is equal to $\alpha_0 - \alpha_1 + \alpha_2$ as obtained after subdividing all faces into triangles. It follows that for any polyhedron in 3 dimensions,

$$V - E + F = 2 \tag{5.43}$$

which is a classic result in mathematics known as the *Euler relation.*

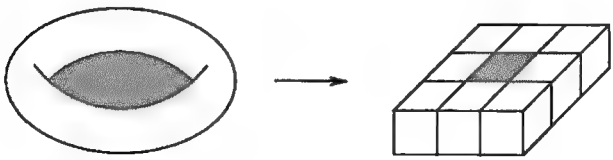


Figure 5.7: "Triangulation" of a 2-torus (by squares).

Example: Another convenient case to study is the two-dimensional torus $S^1 \otimes S^1 = T^2$. This can be easily "triangulated" by squares, as shown in Fig. 5.7. From the discussion above, it is possible to treat the squares as 2-simplices even though the latter should really be triangles. Studying the figure (including the

part of it that is not visible!), we find:

$$\begin{aligned} \alpha_0 &= 32 \\ \alpha_1 &= 24 + 24 + 12 + 4 = 64 \\ \alpha_2 &= 8 + 8 + 12 + 4 = 32 \end{aligned} \tag{5.44}$$

and hence

$$\chi(T^2) = 32 - 64 + 32 = 0 \tag{5.45}$$

So we learn that the 2-torus has vanishing Euler characteristic.

Exercise: Consider a general 2-dimensional manifold, which has the form of a surface with many "handles" (the sphere and torus have 0 and 1 handle respectively). By breaking it up into tori and triangulating each torus, show that a surface Σ_g with g handles has Euler characteristic

$$\chi(\Sigma_g) = 2 - 2g \tag{5.46}$$

5.2 De Rham Cohomology

We now introduce a method to characterise the topology of a manifold in terms of the properties of *differential forms*. Although individual differential forms depend on the local properties of a manifold, it turns out that one can define a structure very analogous to the homology discussed above, on the space of forms, and this reproduces the same topological invariants (the Betti numbers) that we obtained through simplicial homology.

Recall that the *exterior algebra* $\cup_{p=0}^{n} \wedge^p(M)$ is the space of all p-forms on $M, 0 \leq p \leq n$ (where n is the dimension of the manifold). Now each p-form is mapped onto some $(p+1)$-form (possibly zero) by the exterior derivative d, which satisfies $d^2 = 0$. Thus we have a sequence of maps:

$$\ldots \overset{d_{p-2}}{\rightarrow} \wedge^{p-1}(M) \overset{d_{p-1}}{\rightarrow} \wedge^p(M) \overset{d_p}{\rightarrow} \wedge^{p+1}(M) \overset{d_{p+1}}{\rightarrow} \ldots \tag{5.47}$$

where d_p is the d operator on p-forms, viewed as a group homomorphism between the vector spaces $\wedge^p$ and $\wedge^{p+1}$.

In perfect analogy with the definitions of cycles and boundaries, we now define two special kinds of forms:

(i) If $d\omega_p = 0$ then the form ω_p is said to be *closed*.

(ii) If $\omega_p = d\omega_{p-1}$ for some form ω_{p-1}, then ω_p is said to be *exact*.

Moreover, the p-forms on M form an Abelian group, indeed a real vector space $\wedge^p(M)$. Closed and exact forms span subspaces of this vector space, or in other words they are Abelian subgroups of this Abelian group. We define

$$\begin{aligned} Z^p(M) &\equiv \{\, \omega_p \in \wedge^p(M) \mid d\omega_p = 0 \,\} \\ B^p(M) &\equiv \{\, \omega_p \in \wedge^p(M) \mid \omega_p = d\omega_{p-1} \,\} \end{aligned} \tag{5.48}$$

Thus $Z^p(M)$ and $B^p(M)$ are, respectively, the groups formed by the closed and the exact forms on M. All exact forms are closed, by virtue of $d^2 = 0$, therefore $B^p(M) \subset Z^p(M) \subset \wedge^p(M)$ (as vector spaces, hence as groups).

A close analogy is apparent, in structure and in notation, between these concepts for the differential forms and the corresponding notions for simplicial complexes. Some important differences should be noted. While for simplices we started by taking linear combinations with integer coefficients, for forms we are forced at the outset to allow linear combinations with real coefficients, since forms are intrinsically real-valued. Another point is that simplices were themselves an auxiliary construction allowing us to define topological invariants for a manifold, while in the case of forms we work directly on the manifold itself. An important point about notation is that for differential forms, the label p on the groups of closed and exact forms is always written as a superscript, by convention, whereas for p-chains it is a subscript.

After all this, the following definition should come as no surprise.

Definition: The p-th *de Rham cohomology group* $H^p(M)$ of the manifold M, is the quotient of the group of closed forms by the group of exact forms:

$$H^p(M) = Z^p(M)/B^p(M) \tag{5.49}$$

Thus, the elements of $H^p(M)$ are the *equivalence classes* of closed forms which differ by exact forms. Again we clearly have $H^p(M) = 0$ for $p > n$ (where n is the dimension of the manifold) since by antisymmetry there are no forms of dimension larger than that of the manifold. So there is a finite number of de Rham cohomology groups for each manifold, one for each integer p between 0 and n.

As we noted above, forms are intrinsically defined with real (rather than integer) coefficients, so this is a *real* cohomology, and each group has the form:

$$H^p(M) = \mathbb{R} \oplus \ldots \oplus \mathbb{R} \tag{5.50}$$

Suppose now that the manifold M has been smoothly triangulated. Thus there is a homeomorphism between the whole of M and a polyhedron. Pick an open p-simplex σ^p in the polyhedron. This will be homeomorphic to an open submanifold of M, so we can integrate a p-form ω_p over this set (using techniques described in the previous chapter) and call it the *integral of* ω_p *over the simplex* σ^p. Thus we have defined $\int_{\sigma^p} \omega_p$.

This is easily generalized to the integration of forms over chains. Given a p-chain $C_p = \sum_i n_i \sigma_i^p$, we define

$$\int_{C_p} \omega_p = \sum_i n_i \int_{\sigma_i^p} \omega_p \tag{5.51}$$

An obvious generalization of Stokes' theorem tells us that

$$\int_C d\omega = \int_{\partial C} \omega \tag{5.52}$$

where ω is a p-form and C is a $(p+1)$-chain.

In this way we have defined a *pairing* of forms ω and chains C:

$$\langle \omega, C \rangle = \int_C \omega \tag{5.53}$$

where $\omega \in \wedge^p(M)$ and $C \in C_p(K)$. Stokes theorem says that

$$\langle d\omega, C \rangle = \langle \omega, \partial C \rangle. \tag{5.54}$$

Thus we have obtained the fundamental result that the boundary operator ∂ and the exterior derivative d are *dual* to each other. We can use this pairing to associate the de Rham cohomology classes and the simplical homology classes (with real coefficients) to each other.

Let $[\omega]$ be an element of $H^p(M)$ (the square bracket is used to denote an equivalence class). The class $[\omega]$ contains some particular closed form ω satisfying $d\omega = 0$, along with all other forms which differ from ω by an exact form df, where f is an arbitrary $(p-1)$ form. Thus

$$[\omega] = \{\, \omega, \omega + df_1, \omega + df_2, \dots \,\} \tag{5.55}$$

where f_i are all the elements of $\wedge^{p-1}(M)$. This is what a particular de Rham cohomology class looks like.

Next, let $[C]$ be an element of $H_p(M)$, namely, a homology class. Thus $[C]$ contains a given cycle C with $\partial C = 0$, and all others which differ by the addition of boundaries $\partial\phi$, where ϕ is any $(p+1)$-chain. Thus,

$$[C] = \{\, C, C + \partial\phi_1, C + \partial\phi_2, \dots \,\} \tag{5.56}$$

where ϕ_i are all the elements of $C_{p+1}(M)$.

Now the question is whether the pairing $\langle \omega, C \rangle = \int_C \omega$ of forms and chains *induces* a pairing $[\omega]$, $[C]$ of de Rham cohomology classes and simplicial homology classes. In other words, if we define

$$\langle [\omega], [C] \rangle = \langle \omega, C \rangle, \tag{5.57}$$

where ω is any element of $[\omega]$, and C any element of $[C]$, is this independent of the *representatives* $\omega \in [\omega]$, $C \in [C]$ that we chose? This will be true if and only if

$$\langle \omega + df, C + \partial\phi \rangle = \langle \omega, C \rangle \tag{5.58}$$

where f is a $(p-1)$-form and ϕ is a $(p+1)$-chain, both arbitrary.

The desired result is true by linearity and Stokes's theorem:

$$\begin{aligned} \langle \omega + df, C + \partial\phi \rangle &= \langle \omega, C \rangle + \langle \omega, \partial\phi \rangle + \langle df, C \rangle + \langle df, \partial\phi \rangle \\ &= \langle \omega, C \rangle + \langle d\omega, \phi \rangle + \langle f, \partial C \rangle + \langle d^2 f, \phi \rangle \\ &= \langle \omega, C \rangle \end{aligned} \tag{5.59}$$

since $d\omega = 0$, $\partial C = 0$ and $d^2 f = 0$.

Thus we have exhibited the duality between de Rham cohomology and (real) simplical homology for smoothly triangulated manifolds. This is the content of de Rham's theorem. (This actually requires many technical points to be demonstrated in addition to the above simple calculation, but we omit these here.) Thus,

$$H_p(M, \mathbb{R}) \sim H^p(M, \mathbb{R}). \tag{5.60}$$

where we recall that the left hand side, with p as a subscript, denotes the pth simplicial homology group, while the right hand side, with p as a superscript, denotes the pth de Rham cohomology group.

It may seem a bit surprising that the de Rham groups, defined in terms of intrinsically *local* C^∞ forms, capture only global information about the topology of a manifold. In fact the following lemma tells us this must be so.

Poincare's lemma: Every closed form is *locally exact*. In other words, given ω such that $d\omega = 0$, on any finite patch of the manifold we can find f such that $\omega = df$.

Therefore the construction of closed forms modulo exact forms contains no local information. Whatever information it does carry must therefore be global, concerning the topology of the space.

Example: If the whole manifold is homeomorphic to an open disc in $\mathbb{R}^n$, then it follows that all closed forms are exact, and $H^p(M) = 0$ for $p > 0$. In electrodynamics (which we usually do on $\mathbb{R}^3$), $dF = 0$ implies $F = dA$, so there always exists a gauge potential. If there is a pointlike magnetic monopole, then fields are singular at its location so we must work in the modified space $\mathbb{R}^3 - \{0\}$ where the location of the monopole has been removed. Now one can still have $dF = 0$, but for F corresponding to a monopole field there is no globally defined A for which $F = dA$. Thus one ends up either with a Dirac string singularity (by which the space is changed from $R^3 - \{0\}$ to $R^3 - \{\text{infinite half-line}\}$ and then again $F = dA$, or else we must allow for multi-valued gauge potentials. Several physical effects, such as the Bohm-Aharonov effect, follow from similar considerations.

5.3 Harmonic Forms and de Rham Cohomology

We have seen that in de Rham cohomology, an element of $H^p(M)$ is an equivalence class of closed forms which differ by exact forms. We may use the Hodge decomposition theorem as a convenience to uniquely specify a particular element of each class. Recall that according to this theorem (valid on *compact* manifolds),

$$\omega_p = d\alpha + \delta\beta + \gamma \tag{5.61}$$

where $\Delta\gamma_p = 0$, and this decomposition is unique. Now suppose ω is closed, so $d\omega = 0$. Then

$$d(d\alpha + \delta\beta + \gamma) = 0 \tag{5.62}$$

Since $d^2\alpha = 0$, and also $d\gamma = 0$ (every harmonic form is necessarily closed), we have

$$d(\delta\beta) = 0 \tag{5.63}$$

This implies $\langle\beta, d\delta\beta\rangle = \langle\delta\beta, \delta\beta\rangle = 0$. which in turn implies $\delta\beta = 0$. So for a closed form ω,

$$\omega = d\alpha + \gamma \tag{5.64}$$

with $\Delta\gamma = 0$. Thus any closed form is equal to a harmonic form plus an exact form.

This gives a convenient way to label any element of $H^p(M)$. Among all the elements in a cohomology class, *only one* will be harmonic. Then the set of classes which form the cohomology group H^p is equivalent to the set of distinct harmonic p-forms on M, denoted $\mathrm{Harm}^p(M)$. So we can say that

$$H^p(M, \mathbb{R}) \sim \mathrm{Harm}^p(M, \mathbb{R}) \tag{5.65}$$

This result is particularly useful for the following reason. The harmonic p-forms are just the p-forms which solve Laplace's equation,

$$\Delta\omega = 0, \tag{5.66}$$

so they span the kernel of the differential operator Δ. Now Δ is self-adjoint and positive semi-definite, and it is a theorem that such an operator on a compact manifold has a finite-dimensional kernel. So there are finitely many solutions to the above equation. This shows that $\mathrm{Harm}^p(M)$ and hence $H^p(M)$ are *finite-dimensional* on a compact manifold.

Chapter 6

Fibre Bundles

6.1 The Concept of a Fibre Bundle

A fibre bundle is a kind of topological space with additional structure on it. Its main property is that locally, but not always globally, it looks like the *product* of two distinct spaces, one of which is to be thought of as a "fibre" above the other, which is the "base". This concept is of great importance in physics, where the "base" is usually the spacetime manifold, while the fibre represents dynamical variables. Some subtle and important physical phenomena can be conveniently represented in the language of fibre bundles, where they appear more natural and can be associated with topological properties of the bundle.

Let us first discuss the concept of *product space* and *product manifold*. The product of two sets S_1, S_2 is the set

$$S_1 \otimes S_2 = \{\ (x_1, x_2) \mid x_1 \in S_1, x_2 \in S_2\ \} \tag{6.1}$$

If S_1 and S_2 are topological spaces, then let $U_1 = \{u_i\}$, $U_2 = \{u_j\}$ be their respective topologies. We may define a topology, called the *product topology*, on $S_1 \otimes S_2$:

$$U_{S_1 \otimes S_2} = \{u_i \times u_j\},\ u_i \in U_1,\ v_j \in U_2 \tag{6.2}$$

It is easy to check that this defines a topology on the set $S_1 \otimes S_2$. For example, the usual topology generated by open intervals on R leads to a product topology on $\mathbb{R}^2 = \mathbb{R} \otimes \mathbb{R}$ where the open sets are open rectangles:

$$\{\ (x, y) \in R^2 \mid a < x < b,\ c < y < d\ \} \tag{6.3}$$

This is equivalent to the usual topology of open disks on $\mathbb{R}^2$.

If the spaces S_1 and S_2 are also differentiable manifolds, then so is their product. To prove this, we construct an atlas as follows. S_1 has an atlas (U_α, ϕ_α), while S_2 has an atlas (V_β, ψ_β). Then the product manifold $S_1 \otimes S_2$ has an atlas $(U_\alpha \otimes V_\beta, \phi_\alpha \otimes \psi_\beta)$ where $U_\alpha \otimes V_\beta$ are the usual products of sets,

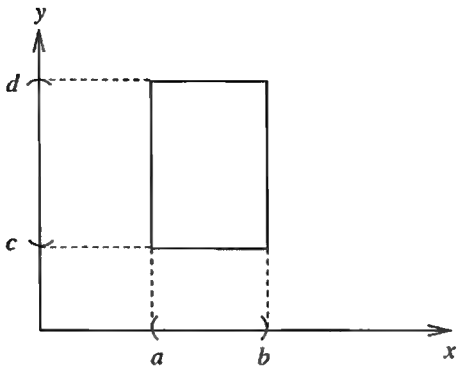


Figure 6.1: The product topology on R^2.

while $\phi_\alpha \otimes \psi_\beta$ are the homeomorphisms:

$$\phi_\alpha \otimes \psi_\beta : \; U_\alpha \otimes V_\beta \to \phi_\alpha(U_\alpha) \otimes \psi_\beta(V_\beta) \tag{6.4}$$

Clearly, if S_1 and S_2 have dimensions m_1, m_2, then $S_1 \otimes S_2$ has dimension $m_1 + m_2$.

An example of a product space is the cylinder, $S^1 \otimes [0,1]$. Since $[0,1]$ is a manifold with boundary, so is $S^1 \otimes [0,1]$. This space is illustrated in Fig. 6.2.

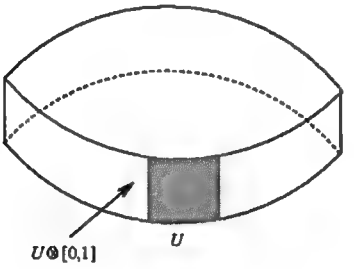


Figure 6.2: The cylinder, $S^1 \otimes [0,1]$. The shaded region is $U \times [0,1]$ where U is an open interval in S^1.

Consider now the Möbius strip (Fig. 6.3), a well-known object that one can construct by folding a strip around and gluing the two ends after giving one of them a twist. It is definitely not a product manifold. Yet, if we take some open set U in S^1, then the shaded region of the Möbius strip does resemble the direct product space $U \otimes [0,1]$. Thus, we may say that the Möbius strip *locally* looks like a direct product manifold.

Let us consider another simple example of a space which locally looks like a direct product. Take the infinite helix, or spiral, and compare it with the direct

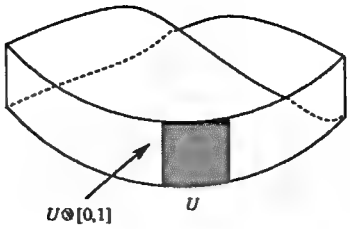


Figure 6.3: The Möbius strip, which locally looks like a product space.

product space $S^1 \otimes \mathbb{Z}$.

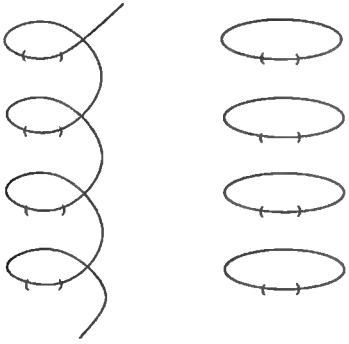


Figure 6.4: The infinite helix and the space $S^1 \otimes Z$.

We can define the helix, illustrated in the first picture in Fig. 6.4, as a subset of 3-dimensional space. It is parametrised by $(\cos 2\pi t, \sin 2\pi t, t)$, $t \in \mathbb{R}$. On the other hand, the space $S^1 \otimes \mathbb{Z}$, illustrated in the second picture in Fig. 6.4, can be parametrised as $(\cos 2\pi t, \sin 2\pi t, n)$, $t \in \mathbb{R}, n \in \mathbb{Z}$ and is quite different – it is an infinite collection of circles, labelled by integers.

Consider now the open sets marked on both diagrams. Clearly they are homeomorphic to each other, and each is homeomorphic to the product space (open interval) $\times \mathbb{Z}$. But the whole helix is not homeomorphic to $S^1 \otimes \mathbb{Z}$! In fact, it is a connected space while $S^1 \otimes \mathbb{Z}$ is not. We will see below that the two spaces in Fig. 6.4 are distinct fibre bundles, but with the same base space and fibre.

Let us restate the situation above somewhat differently. First of all, the circle S^1 can be thought of as the quotient space $\mathbb{R}/\mathbb{Z}$ (the space obtained by identifying all sets of real numbers differing by $2\pi n$). Thus the second space in the figure is homeomorphic to $(\mathbb{R}/\mathbb{Z}) \otimes \mathbb{Z}$. On the other hand, the first one (the helix) is simply homeomorphic to $\mathbb{R}$, and this is certainly not the same as $(\mathbb{R}/\mathbb{Z}) \otimes \mathbb{Z}$, despite the fact that they have homeomorphic subsets. This is why

we say the two spaces are locally, but not globally, equivalent.

Definition: A *fibre bundle* is a space E which is locally like the product of two spaces X and F and possesses the following properties:

(i) E is a topological space, called the *total space.*
(ii) X is a topological space, called the *base space*, and there is a surjective, continuous map $\pi : E \to X$ called the *projection map.*
(iii) F is a topological space, called the *fibre.*
(iv) There is a group G of homeomorphisms of F onto itself, called the *structure group* of the fibre bundle.
(v) There is a collection ϕ of homeomorphisms:

$$\phi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \otimes F \tag{6.5}$$

such that if (x, f) is a point of $U_\alpha \otimes F$ then

$$\pi\phi_\alpha^{-1}(x, f) = x \in X \tag{6.6}$$

(U_α is any open set of the base X).

We now explain all these properties in some detail, and then illustrate them through examples. Let us first discuss the projection map. This simply associates to each point on the fibre bundle a corresponding point on the base space X. This tells us that every point in the bundle E "lies above" a unique point in the base X. π is generally a many-to-one mapping, and the inverse image $\pi^{-1}(x)$ for a given $x \in X$ is the set of all points in E above the point x.

Next, the homeomorphisms ϕ_α give a specific way of regarding the points in E "above" the open set $U_\alpha \subset X$ as being points in the product space $U_\alpha \otimes F$. This is achieved by homeomorphically mapping $\pi^{-1}(U_\alpha)$ to $U_\alpha \otimes F$.

The role of the group G appears when we consider points in X which lie in the intersection of two open sets U_α and U_β. We have the homeomorphisms:

$$\begin{aligned} \phi_\alpha &: \pi^{-1}(U_\alpha \cap U_\beta) \to (U_\alpha \cap U_\beta) \otimes F \\ \phi_\beta &: \pi^{-1}(U_\alpha \cap U_\beta) \to (U_\alpha \cap U_\beta) \otimes F \end{aligned} \tag{6.7}$$

Combining these, we find homeomorphisms:

$$\phi_\alpha \cdot \phi_\beta^{-1} : (U_\alpha \cap U_\beta) \otimes F \to (U_\alpha \cap U_\beta) \otimes F \tag{6.8}$$

Now for each *fixed* $x \in (U_\alpha \cap U_\beta)$, these can be regarded as homeomorphisms,

$$g_{\alpha\beta} = \phi_\alpha \cdot \phi_\beta^{-1} : F \to F \tag{6.9}$$

We require that these homeomorphims form a *group* of transformations of the fibre F; this is called the structure group G.

If the topological spaces are actually differentiable manifolds, then we get (with a few assumptions about differentiability) a *differentiable fibre bundle.*

Let us see how this works in the example of the helix defined earlier. As a topological space, the total space E is the real line $\mathbb{R}$. Chose the base space X to be the circle S^1, and the fibre F to be the set of integers $\mathbb{Z}$. The projection map just takes each point on the helix to the corresponding point on the circle:

$$\pi : (\cos 2\pi t, \sin 2\pi t, t) \rightarrow (\cos 2\pi t, \sin 2\pi t) \in S^1. \tag{6.10}$$

This is a surjective, continuous map $E \rightarrow X$, and $(\cos 2\pi t, \sin 2\pi t)$ parametrises points in S^1, with t being a parameter that ranges from 0 to 1 along the circle.

Given an open interval $U_\alpha = \{\, t \,|\, a < t < b \,\}$ of S^1, $\pi^{-1}(U_\alpha)$ is the set of all points $(\cos 2\pi t, \sin 2\pi t, t + n)$, where $t \in U_\alpha, n \in \mathbb{Z}$ (note that t now has a restricted range, unlike the parameter t introduced earlier while defining the helix. This is because t now represents a point in an open set of the circle S^1). Define the homeomorphism:

$$\phi_\alpha : \pi^{-1}(U_\alpha) \rightarrow U_\alpha \otimes \mathbb{Z} \tag{6.11}$$

by

$$\phi_\alpha\Big((\cos 2\pi t, \sin 2\pi t, t + n)\Big) = (\cos 2\pi t, \sin 2\pi t, n) \tag{6.12}$$

This explicitly shows that the inverse image, under the projection map, of an open set in the base space, is homeomorphic to the direct product of the open set with the fibre.

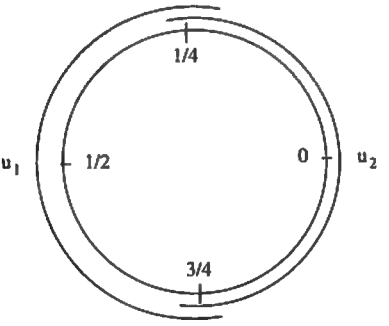


Figure 6.5: Two overlapping open sets U_1, U_2 on the circle S^1.

To complete this example, consider two overlapping open sets of S^1 defined by:

$$\begin{aligned} U_1 &: \frac{1}{4} - \epsilon < t < \frac{3}{4} + \epsilon \\ U_2 &: \frac{3}{4} - \epsilon < t < \frac{5}{4} + \epsilon \end{aligned} \tag{6.13}$$

Then $\phi_i(U_1 \cap U_2), \phi_2(U_1 \cap U_2)$ are given on the upper and lower overlapping regions as follows.

Lower overlap: $\frac{3}{4} - \epsilon < t < \frac{3}{4} + \epsilon$.

$$\begin{aligned} \phi_1(\cos 2\pi t, \sin 2\pi t, t+n) &= (\cos 2\pi t, \sin 2\pi t, n) \\ \phi_2(\cos 2\pi t, \sin 2\pi, t+n) &= (\cos 2\pi t, \sin 2\pi t, n) \end{aligned} \tag{6.14}$$

Clearly the homeomorphism $\phi_2 \cdot \phi_1^{-1}$ is just the identity on this overlap.

The upper overlap is more subtle, since with respect to the patch U_1 the region is $\frac{1}{4} - \epsilon < t < \frac{1}{4} + \epsilon$, while with respect to U_2 it is $\frac{5}{4} - \epsilon < t < \frac{5}{4} + \epsilon$. It is convenient to make the two ranges coincide, in which case n gets replaced by $n-1$ in the image of ϕ_2. Then,

Upper overlap: $\frac{1}{4} - \epsilon < t < \frac{1}{4} + \epsilon$:

$$\begin{aligned} \phi_1(\cos 2\pi t, \sin 2\pi t, t+n) &= (\cos 2\pi t, \sin 2\pi t, n) \\ \phi_2(\cos, 2\pi t, \sin 2\pi t, t+n) &= (\cos 2\pi t, \sin 2\pi t, n-1) \end{aligned} \tag{6.15}$$

Thus

$$\phi_2 \cdot \phi_1^{-1} : \ (U_1 \cap U_2) \otimes \mathbb{Z} \to (U_1 \cap U_2) \otimes \mathbb{Z} \tag{6.16}$$

is given by

$$\phi_2 \cdot \phi_1^{-1}\Big((\cos 2\pi t, \sin 2\pi t, n)\Big) \ \to \ \Big((\cos 2\pi t, \sin 2\pi t, n-1)\Big) \tag{6.17}$$

For fixed t, this can be thought of as the map

$$\phi_2 \cdot \phi_1^{-1} : \mathbb{Z} \to \mathbb{Z} \tag{6.18}$$

defined by

$$\phi_2 \cdot \phi_1^{-1}(n) = n - 1 \tag{6.19}$$

This map, under composition with itself and its inverse, generates the group $G = \mathbb{Z}$ of all integers under addition. Thus the structure group of the bundle in this example is $\mathbb{Z}$.

In this way we have identified all the ingredients in our definition of fibre bundle, in the simple example of the helix. The one special property of this example is that the structure group G and the fibre F are the *same*, namely, $\mathbb{Z}$. We will see later that this example corresponds to a *principal bundle.*

We close this section with a rather obvious definition:

Definition: A fibre bundle with total space E, base space X and fibre F is called a *trivial bundle* if $E = X \otimes F$.

In this case the homeomorphisms $\phi_j \cdot \phi_i^{-1}$ are all equal to the identity and the structure group is trivial.

6.2 Tangent and Cotangent Bundles

In a previous chapter we have defined the tangent space to a manifold M at a point p, denoted $T_p(M)$. This space has the structure of an n-dimensional vector space, where n is the dimension of the manifold. Let us now use it to define a particular fibre bundle over the base space $X = M$. The fibre is $F = T_p(M) \sim \mathbb{R}^n$, and the total space is

$$E = T(M) = \cup_{p \in M} T_p(M) \tag{6.20}$$

(It is always true that the total space of a bundle is the union of the fibres above each point.) This is called the *tangent bundle* of the manifold M.

The projection $\pi : E \to X$ is easy to visualise in this case. It is the map $\pi : T(M) \to M$ defined by $\pi(V \in T_p(M)) = p$. So it associates the tangent space at a point p of the manifold, with the point p itself.

Next we must specify a homeomorphism:

$$\phi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \otimes \mathbb{R}^n \tag{6.21}$$

where U_α is an open set of M. This is provided by the local coordinates on U_α. If $p \in U_\alpha$ and its coordinates are $x^i(p) \in \mathbb{R}^n$, then a tangent vector $V_p \in T_p(M)$ is an element of $\pi^{-1}(U_\alpha)$, and it can be represented as

$$V_p = a^i(x(p)) \frac{\partial}{\partial x^i} \tag{6.22}$$

This defines a map $V \to \Big(p, a^i(x(p))\Big)$ which is the desired homeomorphism from $\pi^{-1}(U_\alpha)$ to $U_\alpha \otimes \mathbb{R}^n$.

Now consider two different coordinate systems x^i and y^i in patches U_α and U_β respectively. In the overlap $U_\alpha \cup U_\beta$ of the two patches, we have the two homeomorphisms

$$\begin{aligned} \phi_\alpha &: V_p \to (p, a^i(p)), \qquad \text{in coordinates } x^i \\ \phi_\beta &: V_p \to (p, b^i(p)), \qquad \text{in coordinates } y^i \end{aligned} \tag{6.23}$$

and $\phi_\beta \cdot \phi_\alpha^{-1}$ relates $a^i(p)$ and $b^i(p)$ by

$$\phi_\beta \cdot \phi_\alpha^{-1} : a^i(p) \to b^i(p) = \left.\frac{\partial y^i}{\partial x^j}\right|_p a^j(p) \tag{6.24}$$

Thus the structure group G consists of all real nonsingular $n \times n$ matrices, in other words $G = GL(n, \mathbb{R})$.

One can similarly construct the *cotangent bundle* $\cup_p T_p^*(M)$, and the *tensor bundles*

$$\cup_p \Big[T_p(M) \otimes T_p(M) \otimes \ldots \otimes T_p^*(M) \otimes T_p^*(M) \otimes \ldots T_p^*(M) \Big] \tag{6.25}$$

All these have structure group $G = GL(n, \mathbb{R})$, although the fibres are different in each case.

6.3 Vector Bundles and Principal Bundles

In many of the above examples, the fibre F is a vector space. Such fibre bundles are special and have a name:

Definition: A *vector bundle* is a fibre bundle whose fibre F is a vector space.

As one more example related to manifolds, consider the orthonormal frames at a point. The bundle obtained by taking the orthonormal frames $\{e^a\}$ at each point of M is called the *orthonormal (tangent) frame bundle.* This is *not* a vector bundle! To specify a point in the fibre, namely a frame, one has to specifiy an $O(n)$ rotation relative to some given frame. Thus the fibre is isomorphic to the group $O(n)$; it is the group manifold of $O(n)$. The structure group of this bundle is also clearly $O(n)$, since across patches one has to specify an $O(n)$ rotation to match the fibres. So in this case, the fibre is the same as the structure group. This type of bundle is also special and has a name:

Definition: A *principal bundle* is a fibre bundle whose fibre F is the structure group G itself.
The example of a helix over S^1 that we encountered earlier is another principal bundle, this time with fibre and structure group isomorphic to $\mathbb{Z}$, the group of integers under addition.

The last few examples have involved base spaces and fibres which are actually differentiable manifolds. They are therefore known as differentiable fibre bundles.

Definition: A *differentiable fibre bundle* is a fibre bundle (E, X, F, π, ϕ, G) for which:
(i) The base space X is an n-dimensional differentiable manifold.
(ii) The fibre F is an m-dimensional differentiable manifold.
(iii) The total space E is an $(m+n)$-dim. differentiable manifold.
(iv) The projection $\pi : F \to X$ is a C^∞ map, of rank n everywhere.
(v) ϕ is a collection of diffeomorphisms (rather than just homeomorphisms).
(vi) G is a Lie group which acts differentiably and effectively. (Thus the map $g : F \to F$ with $g \in G$ is a C^∞ map.)

For a C^∞ manifold, the tangent, cotangent and orthonormal frame bundles are all differentiable fibre bundles.

For a trivial fibre bundle (a direct product of the base space with the fibre), one could define a function on the base space taking values in the fibre. This corresponds to choosing a specific point in the fibre above each point of the base space. For non-trivial bundles we can still do this locally, but not necessarily globally.

Definition: A *local section* (or local cross-section) of a fibre bundle is a continuous injective map $\sigma : U \subset X \to E$ such that $\pi \cdot \sigma(x) = x$.

To define a local section is to continuously pick a point in E "above" every point x in a local region of the base space. The fact that the image is "above" x is imposed by the requirement that if we send the image back down to X by the projection map π, then it lands on the original point x. Thus a local section is precisely a function on an open set of the base space, with values in the fibre.

Definition: A *global section* is a continuous injective map $\sigma : X \to E$ such that $\pi \cdot \sigma(x) = x$.
This differs from a local section in being defined all over the base space, and not just on an open set of it.

Local sections always exist, but global sections may not. It is easy to convince oneself that on the helix there is no global section, but on the Möbius strip there is. The latter is illustrated in Fig. 6.6. After identifying the sides of the rectangle appropriately with the arrows aligned, the wavy line in the figure is a global section of the Möbius strip viewed as a bundle over S^1.

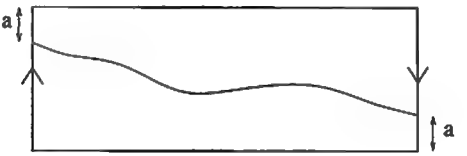


Figure 6.6: A global section of the Möbius strip.

A global section of the tangent bundle is a continuous C^∞ vector field all over the manifold, while a global section of the cotangent bundle is a C^∞ 1-form all over the manifold. It is easy to see that such sections *always* exist. Indeed, since the structure group $G = GL(n, \mathbb{R})$ acts as a linear homogeneous transformation, the zero vector field is invariant, and one can define it everywhere. Similarily the 1-form $\omega = 0$ is a global section of the cotangent bundle.

Whether there exist continuous *everywhere non-zero* sections of the tangent bundle is another story. This depends on the base manifold. For the two sphere S^2, there is a well-known theorem sometimes stated as follows: "You can't comb the hair on a billiard ball." If a 2-sphere were represented as a ball and vector fields as "hair" attached at each point of the ball, then "combing" the hair would amount to flattening this set of vectors of nonzero length to point tangentially to the surface. This would be an everywhere nonzero vector field. In fact it is easy to convince oneself that this is impossible. There is always some point on the ball where the "hair" has no possible direction consistent with continuity. In other words, S^2 admits *no* continuous, everywhere non-zero vector field.

Exercise: Try to find a more precise proof that there is no continuous nowhere-vanishing vector field on S^2.

One would like to have some criteria to decide whether a given bundle is

trivial or not, since it is really in the latter case that they are most interesting. We will quote a few theorems without proof.

Theorem: Given a fibre bundle with base space X and structure group G, if either X or G is *contractible* then the bundle is trivial. (The converse is not not true! A bundle may be trivial even though X and G are non-contractible.)

This theorem has a corollary which is useful even when the spaces are not completely contractible. This says that we can replace G by any subgroup of the same homotopy type, and similarly for X, without changing the topological character of the bundle.

Note that a topological space with the discrete topology is *not* contractible. To be contractible, a space should have the property that the identity map, which sends each element to itself, and the constant map, which sends each element to a fixed element, are homotopic. This is never true with the discrete topology.

We have not yet arrived at any criterion that can tell us whether a given fibre bundle is trivial or not. The following definition will help us find such a criterion.

Definition: Given any bundle E with fibre F and structure group G, we can construct the *associated principal bundle*, denoted $P(E)$, by simply replacing the fibre F with the structure group G.

This is a useful construction because of the following result:

Theorem: The bundles $P(E)$ and E are both trivial if and only if $P(E)$ admits a global section.

Let us use these theorems to study some fibre bundles.

Examples:

(i) The Mobius strip. Here the structure group G is Z_2, whose nontrivial element acts as a twisting on one end of the strip before it is glued to the other end. The associated principal bundle therefore has a fibre that consists of precisely two points. Indeed, this bundle is just the boundary of the Möbius strip.

Clearly the principal bundle has no global section, so the Möbius strip is a non-trivial bundle. It should be recalled that the Möbius strip itself certainly does admit global sections, but that tells us nothing about its non-triviality since it is not a principal bundle. However, by the theorem above, once we have shown that the associated principal bundle is nontrivial then we can be sure the same result also holds for the Möbius strip.

(ii) $T(S^2)$, the tangent bundle of S^2. To show its nontriviality, we will use the fact, referred to above, that there is no everywhere non-zero vector field on S^2.

The associated principal bundle to $T(M)$ is the bundle whose fibres are $GL(n, R)$. This is called the *frame bundle* (note that it is *not* orthonormal!), and is denoted $F(M)$. Now an element of $GL(n, R)$ can be specified by a

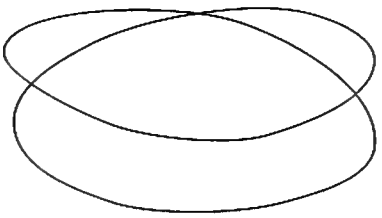


Figure 6.7: The principal bundle associated to the Möbius strip.

collection of n linearly independent, non-zero real n-vectors (these just make up the columns of the matrix). The existence of a global section of the principal bundle associated to $T(M)$ thus reduces to the existence of a collection of n independent vector fields which are everywhere non-zero. But no such vector field exists on S^2, so $F(S^2)$ and hence also $T(S^2)$ are non-trivial bundles.

One can contract $GL(n, R)$ to $O(n)$, from which we get the additional result (using the corollary to a theorem above) that the orthonormal frame bundle is non-trivial on S^2.

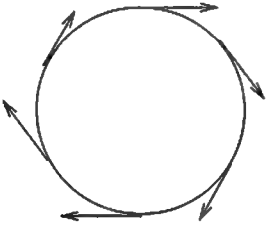


Figure 6.8: A non-zero continuous vector field on S^1.

Definition: A manifold M for which the orthonormal frame bundle has a global section is said to be *parallelisable*.

Thus, the tangent bundle $T(M)$ to a manifold M is a trivial bundle if and only if M is parallelisable.

Example: S^1 is parallelisable. In fact we only need one non-zero continuous tangent vector field, which is depicted in Fig. 6.8. Hence, $T(S^1)$ is trivial, and is in fact the infinite cylinder $S^1 \times R$.

We conclude with a result which is certainly not easy to prove, but is quite remarkable.

Theorem: The only compact manifolds which are parallelizable are the manifolds of all the compact Lie groups and in addition the 7-sphere S^7. (Note

that among the spheres S^n, the only parallelizable ones are S^1, S^3 and S^7. The first two are Lie group manifolds, corresponding to the group $U(1)$ and $SU(2)$ respectively.)

Bibliography

[1] I.M. Singer and J.A. Thorpe: 'Lecture Notes on Elementary Topology and Geometry', *Springer* (1976).

[2] T. Eguchi, P.B. Gilkey and A.J. Hanson, 'Gravitation, Gauge Theories and Differential Geometry', *Phys. Rep 66*, 213 (1980).

[3] C. Nash and S. Sen, 'Topology and Geometry for Physicists', *Academic Press* (1988).

[4] M. Nakahara, 'Geometry, Topology and Physics', *Taylor and Francis* (2nd edition, 2003).

Part II: Group Theory and Structure and Representations of Compact Simple Lie Groups and Algebras

N. Mukunda

Indian Institute of Science (retired)
Bangalore 560 012, India

+Jawaharlal Nehru Fellow

Introduction to Part II

The aim of these notes is to provide the reader with a concise account of basic aspects of groups and group representations, in the form needed for most applications to physical problems. While the emphasis is not so much on completeness of coverage and mathematical rigour as on gaining familiarity with useful techniques, it is hoped that these notes would enable the reader to go further on his or her own in any specialised aspects. The main interest will be in Lie groups and Lie algebras; compact simple Lie algebras; their classification and representation theory; and some specialised topics such as the use of Dynkin diagrams, spinors, etc.

These notes are intentionally written in an informal style, with problems and exercises occasionally inserted to help the reader grasp the points being made.

Chapter 7

Review of Groups and Related Structures

7.1 Definition of a Group

A group G is a set with elements $a, b, \ldots, e, \ldots, g, \ldots$ endowed with the following properties:

(i) A group composition law is given, so that for any two elements of G taken in a definite sequence, a unique product element is determined:

$$a, b \in G \Rightarrow c = ab \in G \tag{7.1}$$

In general the sequence is relevant, so ab and ba may be different.

(ii) This composition is associative, so for any three elements taken in a definite sequence, the product is unambiguous:

$$a, b, c \in G : a(bc) = (ab)c \tag{7.2}$$

(iii) There is a distinguished element $e \in G$, the identity, which obeys:

$$a \in G : ae = ea = a \tag{7.3}$$

(iv) For each $a \in G$, there is a unique inverse element $a^{-1} \in G$, such that

$$a^{-1}a = aa^{-1} = e. \tag{7.4}$$

These laws defining a group are not expressed here in the most economical form. One can show that e is unique; that if one had defined separate left and right inverses, they would have been the same; that for each a, its inverse a^{-1} is unique; etc.

Familiar examples of groups are: S_n, the group of permutations of n objects, with $n!$ distinct elements; the group SO(3) of proper real orthogonal rotations

in three dimensional space; the groups of Lorentz and of Poincaré transformations in three space and one time dimension; the Galilei group relevant to non-relativistic mechanics; the discrete group of translations of a crystal lattice.

Some groups have a finite number of elements (this is then the order of the group); others have a denumerable infinity of elements, and yet others a continuous infinity of elements which can be parametrised by a finite number of real independent continuously varying parameters.

7.2 Conjugate Elements, Equivalence Classes

Given any two elements $a, b \in G$, we say they are conjugate to one another if there is an element $c \in G$ (may be more than one) such that

$$b = cac^{-1} \tag{7.5}$$

This is an equivalence relation, written as $a \sim b$, because the three properties of such a relation do hold:

(i) $a \sim a$: take $c = e$;

(ii) $a \sim b \Rightarrow b \sim a$: use c^{-1} in place of c;

(iii) $a \sim b, b \sim c \Rightarrow a \sim c$

These are, respectively, the identity law, the reflexive and the transitive properties.

The group G thus splits into disjoint conjugacy classes or equivalence classes. To the class of $a \in G$ belong all elements $b \in G$ conjugate to a; elements in different classes are non conjugate; two different classes have no common elements; the identity e forms a class all by itself.

As examples, let us mention the following: in S_n, all elements with a given cycle structure belong to one equivalence class; in SO(3), all rotations by a given angle but about all possible axes form one class.

7.3 Subgroups and Cosets

A subset H in a given group $G, H \subset G$, is a subgroup if its elements obey all the laws of a group, it being understood that the composition law is the one given in G. Taking advantage of the fact that products and inverses of elements in G are already defined, we can express concisely the condition for a subset H to be a subgroup:

$$h_1, h_2 \in H \Rightarrow h_1^{-1} h_2 \in H \tag{7.6}$$

(Verify that this condition is necessary and sufficient for H to be a subgroup).

The cases $H = G$ or $H = \{e\}$ lead to trivial subgroups. Every other subgroup is called a proper or nontrivial subgroup.

Given a subgroup $H \subset G$, two (in general different) equivalence relations can be set up in G using H: one is called left equivalence with respect to H, and the other right equivalence with respect to H. These are to be distinguished from one another, and from equivalence defined by the conjugacy property (which does not involve any subgroup). For any two elements $a, b \in G$,

$$a \text{ and } b \text{ are left equivalent with respect to } H \Leftrightarrow a^{-1}b \in H \Leftrightarrow b = ah \text{ for some } h \in H \quad (7.7)$$

(Verify that all the laws of an equivalence relation are obeyed). The corresponding equivalence classes are called "the left cosets of H in G". The left coset containing $a \in G$ is a subset written as

$$aH = \{ah \mid h \in H\} \quad (7.8)$$

A left coset is determined by any of its elements, any two left cosets are identical or disjoint, and G is the union of left cosets. The left coset containing e is H itself; every other one does not contain e and so is not a subgroup. Right equivalence is set up similarly:

$$a \text{ and } b \text{ are right equivalent with respect to } H \Leftrightarrow ab^{-1} \in H \Leftrightarrow a = hb \text{ for some } h \in H \quad (7.9)$$

The corresponding equivalence classes are called right cosets of H in G, a typical one being

$$Ha = \{ha \mid h \in H\} \quad (7.10)$$

These too are mutually disjoint, etc.

For a general subgroup $H \subset G$: the set of left cosets and the set of right cosets are two generally different ways of separating G into disjoint subsets. Each coset has the same "number of elements" as H.

7.4 Invariant (Normal) Subgroups, the Factor Group

If H is a subgroup of G, and $g \in G$ is a general element, the set of elements

$$H_g = \{ghg^{-1} | h \in H, g \text{ fixed}\} \equiv gHg^{-1} \quad (7.11)$$

is also a subgroup of G. If $g \in H$, then $H_g = H$; otherwise, in general we would expect $H_g \neq H$.

The subgroup H is said to be an *invariant*, or *normal*, subgroup if $H_g = H$ for every $g \in G$. In other words,

$$H \text{ is an invariant subgroup } \Leftrightarrow h \in H, g \in G \Rightarrow ghg^{-1} \in H. \quad (7.12)$$

For such a subgroup, the break up of G into left and into right cosets coincide, since each left coset is a right coset and conversely:

$$H \text{ invariant} \Leftrightarrow aH = Ha \quad \text{for each } a \in G \tag{7.13}$$

These (common) cosets can now be made into the elements of a new group! Namely, we define the product of two cosets aH and bH to be the coset containing ab:

$$\begin{aligned} \text{Coset composition: } aH \cdot bH &= \{ahbh' \mid h, h' \in H\} \\ &= \{abh \mid h \in H\} \\ &= abH \end{aligned} \tag{7.14}$$

The group obtained in this way is called the *factor group* G/H: its elements are the H-cosets of G, the identity is H itself, etc. (Verify that since H is invariant, the definition (7.14) obeys all the group laws).

This construction cannot be carried out if H is not invariant.

7.5 Abelian Groups, Commutator Subgroup

A group G is said to be *abelian*, or *commutative*, if the product of elements does not depend on their sequence:

$$G \text{ abelian} : ab = ba \text{ for every } a, b \in G. \tag{7.15}$$

If it is not so, we say G is *nonabelian* or *noncommutative*.

The translations in space form an abelian group. The permutation group S_n for $n \geq 3$, the rotation group SO(3), and the Lorentz group are all nonabelian.

For a general group G, we can "measure the extent to which two elements a and b do not commute" by defining their *commutator* to be another element of G:

$$\begin{aligned} q(a, b) &= \text{commutator of } a \text{ and } b \\ &= aba^{-1}b^{-1} \in G \end{aligned} \tag{7.16}$$

(This notion of the commutator of two group elements is closely related to the commutators of operators familiar in quantum mechanics). Clearly we can say:

$$q(a, b) = e \Leftrightarrow ab = ba \Leftrightarrow a \text{ and } b \text{ commute} \tag{7.17}$$

The "function" $q(a, b)$ has the following obvious properties:

$$\begin{aligned} q(a, b)^{-1} &= q(b, a), \\ cq(a, b)c^{-1} &= q(cac^{-1}, cbc^{-1}) \end{aligned} \tag{7.18}$$

These properties allow us to define what can be called the *commutator subgroup* of G, to be denoted as $Q(G,G)$. This notation will become clear later on. We define:

$$\begin{aligned} Q(G,G) &= \text{products of any numbers of commutators of pairs of} \\ &\qquad \text{elements in } G \\ &= \text{collection of elements } q(a_m,b_m)q(a_{m-1},b_{m-1})\dots q(a_2,b_2) \\ &\qquad q(a_1,b_1) \text{ for all } m \text{ and all choices of } a\text{'s and } b\text{'s} \end{aligned} \tag{7.19}$$

We see immediately that, using Eq.(7.18) when necessary:

(i) $Q(G,G)$ is an invariant subgroup of G;

(ii) the factor group $G/Q(G,G)$ is abelian.

Property (ii) utilises the fact that, on account of $ab = q(a,b)ba$, ab and ba are in the same coset. One can extend the argument to say: if $S \subset G$ is an invariant subgroup of G such that $Q(G,G)$ is contained in it, i.e., S is "somewhere in between $Q(G,G)$ and G", then G/S too will be abelian. Conversely, one can also show easily that if S is an invariant subgroup of G such that G/S is abelian, then S must contain $Q(G,G)$. In other words, $Q(G,G)$ is the smallest invariant subgroup in G such that the associated factor group is abelian.

All these properties entitle us to say that the commutator subgroup of a given group captures in an intrinsic way "the extent to which the group is non-abelian". In fact, one has trivially the extreme property

$$G \text{ is abelian } \Leftrightarrow Q(G,G) = \{e\} \tag{7.20}$$

7.6 Solvable, Nilpotent, Semisimple and Simple Groups

We can use the notions of invariant subgroups and of the commutator subgroup to develop several other very useful concepts. Given a group G, form the sequence of groups

$$\begin{aligned} G_1 &= Q(G,G) = \text{commutator subgroup of } G; \\ G_2 &= Q(G_1,G_1) = \text{commutator subgroup of } G_1; \\ &\dots \quad \dots \quad \dots \\ G_{j+1} &= Q(G_j,G_j) = \text{commutator subgroup of } G_j; \end{aligned} \tag{7.21}$$

The series of groups

$$G_0 \equiv G, G_1, G_2, \dots, G_j, G_{j+1}, \dots \tag{7.22}$$

is then such that each group here is a normal subgroup of the previous one, and the corresponding factor group is abelian:

$$G_j/G_{j+1} = \text{abelian}, \; j = 0,1,2,\dots \tag{7.23}$$

We say the group G is *solvable* if for some finite j, G_j becomes trivial:

$$G \text{ is solvable } \Leftrightarrow G_n = \{e\} \text{ for some finite } n \tag{7.24}$$

Solvability is an important generalisation of being abelian. (The name has its origin in properties of certain systems of differential equations). If G is solvable, there is of course a least value of n for which Eq.(7.24) is satisfied, which is then the significant value. If G is abelian, then Eq.(7.24) is obeyed already for $n = 1$, so G is solvable. If G is not solvable, then for no value of n (however large!) does G_n become trivial!

In contrast to the series of successive commutator subgroups (7.22), another interesting series is the following. Given G, we form $Q(G,G)$ to begin with, but now write it as G^1 instead of as G_1. Then we define $G^2, G^3, \ldots$ successively thus:

$$\begin{aligned} G^2 &= Q(G,G^1) = \text{products of any numbers of elements of the form} \\ &\qquad q(a,b) \text{ for } a \in G \text{ and } b \in G^1, \text{ and their inverses;} \\ G^3 &= Q(G,G^2) = \text{similarly defined but with } G^2 \text{ in place of } G^1 \\ &\ldots \\ G^{j+1} &= Q(G,G^j) \end{aligned} \tag{7.25}$$

Each of these is in fact a group, and we have then the series of groups

$$G_0 \equiv G^0 \equiv G, G^1 \equiv G_1 = Q(G,G^0), G^2 = Q(G,G^1), \ldots, G^{j+1} = Q(G,G^j) \tag{7.26}$$

The successive members in this series are *not* formed by the commutator subgroup construction. Nevertheless, it is an instructive exercise to convince oneself that for each j,

(i) G^{j+1} is an invariant subgroup of G^j, and

(ii) G^j/G^{j+1} is abelian.

Now the notation $Q(G,G)$ for the commutator subgroup is clear: it is a special case of the more general object appearing in Eq.(7.25).

We say the group G is *nilpotent* if for some n (and so for some least value of n), G^n is trivial:

$$G \text{ is nilpotent } \Leftrightarrow G^n = \{e\} \text{ for some finite } n \tag{7.27}$$

Nilpotency is a stronger property than solvability:

$$G \text{ is nilpotent } \Rightarrow G \text{ is solvable, but not conversely} \tag{7.28}$$

In our later discussions concerning Lie groups, we shall have to deal with solvable groups to some extent, but not much with nilpotent ones.

Two other definitions are important: these are of *simple* and *semisimple* groups. We say:

$$\begin{gathered} G \text{ is simple} \Leftrightarrow G \text{ has no nontrivial invariant subgroups;} \\ G \text{ is semisimple} \Leftrightarrow G \text{ has no nontrivial abelian invariant subgroups} \end{gathered} \tag{7.29}$$

The four notions – solvable, nilpotent, simple, semisimple – can be related to one another in various ways. On the one hand, as a counterpart to Eq.(7.28), we obviously have:

$$G \text{ is simple} \Rightarrow G \text{ is semisimple, but not conversely} \tag{7.30}$$

On the other hand (leaving aside the case of abelian G when $G_1 = G^1 = \{e\}$), if G is solvable, then $G_1 = G^1 = Q(G,G)$ must be a proper invariant subgroup of G, so G is not simple:

$$G \text{ is solvable} \Rightarrow G \text{ is not simple} \tag{7.31}$$

By the same token, if G is simple, then $G_1 = Q(G,G)$ must coincide with G, so G cannot be solvable:

$$G \text{ is simple } \Rightarrow G \text{ is not solvable} \tag{7.32}$$

So to be simple and to be solvable are mutually exclusive properties; semisimplicity is weaker than the former, and nilpotency is stronger than the latter. Of course a given group G need not have any of these four properties, but in a qualitative way one can say that nilpotent groups and simple groups are of opposite extreme natures. In a "linear" way we can depict the situation thus:

$$\text{Nilpotent} \Rightarrow \text{solvable} — \text{General group} — \text{Semisimple} \Leftarrow \text{simple group}$$

7.7 Relationships Among Groups

Let G and G' be two groups, and let us write $a, b, \ldots, g, \ldots$ and $a', b', \ldots, g', \ldots$ for their elements. Are there ways in which we can compare G and G'? When can we say they are "essentially the same"? When can we say that one of them is a "larger version" or a "smaller version" of the other? These are natural qualitative questions that come to mind, and we can set up precise concepts that help us answer them: homomorphism, isomorphism and automorphism.

A *homomorphism* from G to G' (the order is important!) is a mapping from G to G' obeying:

$$\begin{gathered} \varphi : G \to G' : a \in G \Rightarrow \varphi(a) = a' \in G'; \\ \varphi(a)\varphi(b) = \varphi(ab) \end{gathered} \tag{7.33}$$

For each $a \in G$, of course φ must assign a unique unambiguous image $\varphi(a) \in G'$; then multiplication in G "goes over into" multiplication in G', so the homomorphism property is essentially that

$$\text{product of images} = \text{image of product.}$$

The following are easy and immediate consequences:

$$e = \text{identity of } G : \varphi(e) = e' = \text{identity of } G';$$
$$\varphi(a)^{-1} = \varphi(a^{-1}) \tag{7.34}$$

In general, we must recognise that in a homomorphism,

(i) more than one element in G may have the same image in G', i.e., φ could be many-to-one;

(ii) we may not get all of G' as we allow a in $\varphi(a)$ to run over all of G; so $\varphi(G) \subseteq G'$.

We define next another kind of relationship between groups, then return to a discussion of homomorphisms. A homomorphism is an *isomorphism* if the map $\varphi : G \to G'$ is one-to-one and onto, hence invertible. In that case we have a one-to-one correspondence between elements of G and G', such that the multiplication laws are respected, and we can then really say the two groups are "essentially the same" and cannot be distinguished as groups. We say that they are *isomorphic.*

Now return to the case of a homomorphism $\varphi : G \to G'$. It is easy to check the following:

(i) $\varphi(G) = \{\varphi(a) | a \in G\} =$ subgroup of G';

(ii) the kernel K of the homomorphism, defined as

$$K = \{a \in G | \varphi(a) = e'\}$$

is an invariant subgroup of G;

(iii) The factor group G/K is isomorphic to $\varphi(G)$.

So in the case of a homomorphism $\varphi : G \to G'$, the two groups play different roles and the relationship is non-symmetrical or "one-way". We can say that G is "a larger version" of G'. If for simplicity we assume that $\varphi(G) = G'$, then conversely we can say G' is "a smaller version" of G. These are only meant to be qualitative descriptions intended to help picture a homomorphism. Only if φ is an isomorphism are G and G' "essentially the same".

An isomorphism in which the two groups are the same, $G' = G$, is called an *automorphism.* That is, an automorphism τ of a group G is a one-to-one, onto, hence also invertible, mapping of G onto itself preserving the group composition law:

$$\tau : a \in G \to \tau(a) \in G,$$
$$\tau(a)\tau(b) = \tau(ab),$$

$$\tau^{-1} \text{ exists} \tag{7.35}$$

An example of an automorphism is given by conjugation with any fixed element of G, say g. We define τ_g, indexed by g, as

$$\tau_g(a) = gag^{-1} \text{ for each } a \in G \tag{7.36}$$

It is easy to check that this is indeed an automorphism. Each element a stays within its conjugacy class. Such automorphisms are called *inner* automorphisms. They form a group on their own too, because

$$\tau_{g'} \cdot \tau_g = \tau_{g'g} \tag{7.37}$$

Automorphisms which are not inner are called *outer*. The set of *all* automorphisms of a given group G can be shown to form a group, with the inner ones forming an invariant subgroup.

7.8 Ways to Combine Groups — Direct and Semidirect Products

Let G_1 and G_2 be two groups. Their *direct product* $G_1 \times G_2$ is a group defined in the obvious way with natural group operations:

(i) Elements of $G_1 \times G_2$ are ordered pairs $(a_1, a_2), a_1 \in G_1, a_2 \in G_2$.
(ii) Pairs are composed by the rule

$$(a_1, a_2)(b_1, b_2) = (a_1 b_1, a_2 b_2)$$

(iii) The identity and inverses are given as the pair (e_1, e_2), and $(a_1, a_2)^{-1} = (a_1^{-1}, a_2^{-1})$ respectively. That we do have a group here, and that $G_2 \times G_1$ is isomorphic to $G_1 \times G_2$, are both easy to see. The extension to direct products of more factors, such as $G_1 \times G_2 \times G_3$, is also evident.

A more intricate way of combining G_1 and G_2 to produce a third group is the *semidirect product*. This requires more structure. Since the two groups in the construction play very different roles, we prefer to write H and K rather than G_1 and G_2 for them. To form the semi-direct product $H \times) K$, we need for each $k \in K$ an automorphism τ_k of H such that

$$\tau_{k'} \cdot \tau_k = \tau_{k'k} \tag{7.38}$$

In more detail, for each $k \in K$ we need an onto, invertible map τ_k of H to itself obeying all the following:

$$\begin{aligned} \tau_k(h')\tau_k(h) &= \tau_k(h'h) \quad \text{for} \quad h, h' \in H; \\ \tau_{k'}(\tau_k(h)) &= \tau_{k'k}(h) \quad \text{for} \quad k, k' \in K, h \in H. \end{aligned} \tag{7.39}$$

Then the group $H\times)\ K$ is defined as the set of all ordered pairs (h,k) with $h\in H, k\in K$ obeying the law of composition

$$(h',k')(h,k)=(h'\tau_{k'}(h),k'k) \tag{7.40}$$

The second factors in the pairs "mind their own business", while in the composition of the first factors drawn from H, the automorphisms τ play a part. It is an instructive exercise to exploit the properties (7.39) and convince oneself that the composition law (7.40) is an acceptable one. While the identity of $H\times)\ K$ is the ordered pair (e,e') where e and e' are the identity elements in H and K respectively, inverses are given by

$$(h,k)^{-1}=(\tau_{k^{-1}}(h^{-1}),k^{-1}) \tag{7.41}$$

(here one must use the property $\tau(a)^{-1}=\tau(a^{-1})$ for an automorphism!)

The semidirect product reduces to the direct product when all the automorphisms τ are taken to be the identity. Non-trivial examples are numerous: the Euclidean group $E(3)$, the Poincaré group, the space group of a crystal, etc. One can picture the semidirect product construction as a kind of inverse to the passage to the factor group: indeed, $H\times)\ K$ contains as subgroups (e,K) and (H,e') respectively isomorphic to K and H; (H,e') is an invariant subgroup of $H\times)\ K$, and $H\times)\ K/(H,e')$ is isomorphic to K.

The semidirect product will appear later in the general theory of Lie groups.

7.9 Topological Groups, Lie Groups, Compact Lie Groups

An introduction to topological and differential geometric concepts is given in the accompanying notes in this monograph; we use some of them at this stage.

A *topological group* is a set G which is a group and a topological space at the same time, and the group operations (composition, taking of inverses) are continuous in the sense defined by the topology.

A *Lie group of dimension (or order)* r is a topological group G such that G is also a smooth manifold of (real) dimension r. This means that G is expressible as the union of a certain collection of open sets, each of which is homeomorphic to a connected open set of Euclidean real r-dimensional space. Thus, in each of these open sets in G, we can assign r independent, real coordinates to each group element in a smooth way.

A *compact Lie group of dimension* r is an r-dimensional Lie group G such that as a topological space G is compact. This means that every open cover of G contains a finite subcover.

If G is not compact, it is *non-compact.*

Let $\mathfrak{N}$ be an open set (a neighbourhood) containing the identity in a Lie group G. By $a\mathfrak{N}$ for some $a\in G$ we mean the collection of elements

$$a\mathfrak{N}=\{ag|g\in\mathfrak{N}\subset G\} \tag{7.42}$$

Continuity of group operations ensures that $a\mathfrak{N}$ too is open. Then obviously

$$G = \bigcup_{a \in G} a\mathfrak{N} \tag{7.43}$$

provides us with an open covering of G. If G is compact, then this must contain a finite subcover. So in a compact Lie group G, given any neighbourhood $\mathfrak{N}$ containing e, we can find a finite number of elements $a_1, a_2, \ldots, a_N$ in G such that

$$G = a_1\mathfrak{N} \cup a_2\mathfrak{N} \cup \ldots \cup a_N\mathfrak{N} \tag{7.44}$$

We would naturally expect N to be larger for smaller $\mathfrak{N}$ and vice versa.

A compact Lie group is one for which in an intuitive sense, the volume is finite (however one must define volume precisely!) Examples among groups of importance in high energy physics are SO(3), SU(2), SU(3) etc. On the other hand, the Euclidean, Galilei, Lorentz and Poincaré groups are all non-compact. We will later be largely concerned with compact simple Lie groups; to some extent, in connection with spinors, we shall also deal with the pseudo-orthogonal groups $\mathrm{SO}(n, 1)$, $\mathrm{SO}(p, q)$ which are noncompact.

Exercises for Chapter 7

1. Determine which of the following sets with specified operations are groups, whether abelian or non-abelian:

   (i) All complex numbers under addition.

   (ii) All complex numbers under multiplication.

   (iii) All real three dimensional vectors under addition.

   (iv) All real $n \times n$ matrices under multiplication.

   (v) All real $n \times n$ unimodular matrices under multiplication.

   (vi) All real orthogonal/complex unitary $n \times n$ matrices under multiplication.

2. For the permutation groups S_n, review the following:

   (i) Various notations for group elements.

   (ii) Composition law, identity, inverses.

   (iii) Cycle structure of a permutation.

   (iv) Equivalence classes as determined by cycle structure.

   (v) Expression of any permutation as a product of transpositions.

   (vi) Show that S_2 is abelian, S_n for $n \geq 3$ is nonabelian.

3. For the abelian group of translations in n real dimensions, consisting of translation vectors $\underline{a} = (a_1, a_2, \ldots, a_n)$, show that for any real non singular $n \times n$ matrix S,

$$\underline{a} \rightarrow \underline{a}' = S\underline{a}$$

is an automorphism. Is it inner or outer?

4. The rotation groups in the Euclidean plane, proper and improper, are defined as

$$\mathrm{SO}(2)/\mathrm{O}(2) = \{A = 2 \times 2 \text{ real matrix } | A^T A = 1_{2\times 2}, \quad \det A = 1/\pm 1\}$$

Show that SO(2) is abelian, O(2) nonabelian.

5. Show that from a direct product $G_1 \times G_2$ of two groups, the individual factors can be recovered as invariant subgroups.

6. The Euclidean group E(3) has elements $(R, \underline{a})$ where $R \in \mathrm{O}(3)$ the real orthogonal group, and $\underline{a}$ is a 3-dimensional real translation vector. The composition law is

$$(R', \underline{a}')(R, \underline{a}) = (R'R, \underline{a}' + R'\underline{a})$$

Identify the O(3) subgroup, the translation subgroup T_3, and show that the latter is normal. Find the corresponding cosets and show that E(3) is a semidirect product $T_3 \times)$ O(3). Are O(3) and T_3 unique subgroups in E(3)?

Chapter 8

Review of Group Representations

8.1 Definition of a Representation

We shall throughout be concerned with linear representations on finite dimensional real or complex linear vector spaces. Let a group G be given, and let $\mathcal{V}$ be a (real or complex) linear vector space. We say we have a linear representation of G on $\mathcal{V}$ if for each $g \in G$, we have a non-singular linear transformation $D(g)$ on $\mathcal{V}$ such that:

$$\begin{aligned} &\text{(i) } D(g')D(g) = D(g'g) \text{ for all } g, g' \in G; \\ &\text{(ii) } D(e) = 1, \text{ the unit operator on } \mathcal{V}; \\ &\text{(iii) } D(g)^{-1} = D(g^{-1}) \text{ for all } g \in G \end{aligned} \tag{8.1}$$

The dimension of the representation is the dimension of $\mathcal{V}$; and it is real or complex according to the nature of $\mathcal{V}$. The representation is *faithful* if and only if distinct group elements correspond to distinct linear transformations.

The three conditions above are not stated in a minimal and most economical way, but contain some redundancy. This is desirable in the interest of explicitness, similar to the way a group was defined in Section 7.1.

The use of a basis for $\mathcal{V}$ allows us to represent each operator $D(g)$ by a corresponding matrix. For simplicity of notation the same symbol $D(g)$ will be used for both the abstract linear transformation and its matrix representative in a definite basis. From the context the meaning will always be clear.

If we pick a basis $\{e_j\}, j = 1, 2, \ldots, n$ for $\mathcal{V}$, made up of linearly independent vectors, then the matrix associated with $D(g)$ is obtained as follows:

$$g \in G : D(g)e_j = D(g)_{kj}e_k \tag{8.2}$$

Here a summation on the repeated index k is understood, and the positions of the indices j and k must be carefully noticed. The composition law for the

transformations, Eq.(8.1)(i), then translates into matrix form as

$$D(g')_{jk}D(g)_{kl} = D(g'g)_{jl}, \tag{8.3}$$

and for the identity element we have the unit matrix:

$$D(e)_{jk} = \delta_{jk} \tag{8.4}$$

In this way, for each $g \in G$ we have an $n \times n$ matrix $D(g)$ with elements $D(g)_{jk}$; it is real if $\mathcal{V}$ is real, and may be real or complex if $\mathcal{V}$ is complex. These representation matrices follow, upon multiplication, the group composition law.

We are of course free to replace the basis $\{e_j\}$ by another one, $\{e'_j\}$ say. The effect on the representation matrices is then a similarity transformation. It is an elementary exercise to check that this is so:

$$\begin{aligned} e_j \rightarrow e'_j = S^{-1}_{kj} e_k \Rightarrow e_j = S_{kj} e'_k; \\ D(g)_{jk} \rightarrow D'(g)_{jk} = S_{jl} D(g)_{lm} S^{-1}_{mk}, \\ \text{i.e.} \qquad D'(g) = SD(g)S^{-1} \end{aligned} \tag{8.5}$$

It is of course desirable to think of group representations and their most important properties as far as possible in an intrinsic and basis independent way. It is thus that one gets "to the heart of the matter". However, when expedient, and for practical purposes, one need not hesitate to use specific bases, matrices etc. In Feynman's words, this does not by any means involve a sense of defeat!

8.2 Invariant Subspaces, Reducibility, Decomposability

Let $g \rightarrow D(g)$ be a linear representation of the group G on the (real or complex) linear vector space $\mathcal{V}$. If there is a non-trivial subspace $\mathcal{V}_1 \subset \mathcal{V}$, neither equal to the full space $\mathcal{V}$ nor to the zero space 0, such that

$$x \in \mathcal{V}_1, g \in G \Rightarrow D(g)x \in \mathcal{V}_1, \tag{8.6}$$

we say the subspace $\mathcal{V}_1$ is *invariant* under the given representation, and that the representation itself is *reducible*. If there is no such nontrivial invariant subspace, then the representation is *irreducible*.

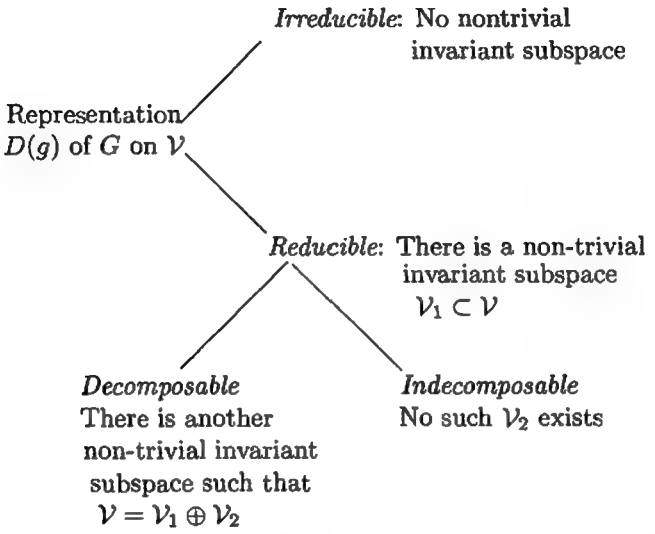
Suppose a given representation is reducible, on account of the existence of the non-trivial invariant subspace $\mathcal{V}_1 \subset \mathcal{V}$. If we can find another nontrivial subspace $\mathcal{V}_2 \subset \mathcal{V}$, such that

(i) $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2$,

(ii) $\mathcal{V}_2$ is also an invariant subspace,

then the given representation is *decomposable*. If such $\mathcal{V}_2$ cannot be found, then the representation is *indecomposable*.

Thus the various possibilities can be depicted diagrammatically thus:



In matrix form, these possibilities can be recognised more graphically. If a basis for $\mathcal{V}$ is made up of a basis for $\mathcal{V}_1$ plus some additional basis vectors, then reducibility tells us that in such a basis the representation matrices have the forms

$$\textbf{Reducible: } D(g) = \begin{pmatrix} D_1(g) & \vdots & B(g) \\ \cdots & \cdots & \cdots \\ 0 & \vdots & D_2(g) \end{pmatrix} \text{ for all } g \in G \qquad (8.7)$$

where $D_1(g)$ and $D_2(g)$ are both non-singular, and in fact both form representations of G. The indecomposable case is when it is impossible to reduce the off diagonal block $B(g)$ to zero; in the decomposable case, we are able to choose the additional basis vectors to span the invariant subspace $\mathcal{V}_2$, so $B(g)$ does vanish:

$$\textbf{Decomposable: } D(g) = \begin{pmatrix} D_1(g) & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & D_2(g) \end{pmatrix} \text{ for all } g \in G \qquad (8.8)$$

Of course, in the irreducible case, even the form (8.7) cannot be achieved for all elements g.

In the reducible decomposable case, the original representation $D(\cdot)$ on $\mathcal{V}$ is really obtained by "stacking" or putting together the two representations $D_1(\cdot)$ on $\mathcal{V}_1$ and $D_2(\cdot)$ on $\mathcal{V}_2$. We call this the process of forming the *direct sum* of

these two representations. Quite generally, let representations $D_1(\cdot)$ and $D_2(\cdot)$ on (both real or both complex!) linear vector spaces $\mathcal{V}_1$ and $\mathcal{V}_2$ be given. First form the direct sum vector space $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2$, so that

$$x \in \mathcal{V} \Rightarrow x = x_1 + x_2 \text{ uniquely,} \quad x_1 \in \mathcal{V}_1, x_2 \in \mathcal{V}_2 \tag{8.9}$$

Then the direct sum of the two representations $D_1(\cdot)$ and $D_2(\cdot)$ is the representation $D(\cdot)$ on $\mathcal{V}$ with the action

$$D(g)x = D_1(g)x_1 + D_2(g)x_2 \tag{8.10}$$

The direct sum construction can obviously be extended to three or more summands.

Go back now to a given reducible decomposable representation $D(\cdot)$ on $\mathcal{V}$. Now we raise the same question of reducibility for $D_1(\cdot)$ and $D_2(\cdot)$. If each of these is either irreducible or reducible decomposable, we can keep posing the same question for their "parts", and so on. If we can in this way break up $D(\cdot)$ into finally irreducible pieces, we then say that $D(\cdot)$ is a *fully reducible* representation. That is to say, a reducible representation $D(\cdot)$ on $\mathcal{V}$ is fully reducible if we are able to express $\mathcal{V}$ as the direct sum of invariant subspaces,

$$\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \ldots \mathcal{V}_N \tag{8.11}$$

and $D(\cdot)$ restricted to each of these is irreducible. For all the groups we shall deal with, every representation is either irreducible or fully reducible. In particular this is true for all finite groups, and also for any simple group.

8.3 Equivalence of Representations, Schur's Lemma

For a given group G, suppose $D(\cdot)$ and $D'(\cdot)$ are two representations on the spaces $\mathcal{V}$ and $\mathcal{V}'$, both being real or both complex. We say that these representations are equivalent if and only if there is a one-to-one, onto, hence invertible linear mapping $T : \mathcal{V} \to \mathcal{V}'$ such that

$$D'(g) = TD(g)T^{-1} \text{ for all } g \in G \tag{8.12}$$

Clearly, $\mathcal{V}$ and $\mathcal{V}'$ must then have the same dimension. In practice, when we deal with equivalent representations, the spaces $\mathcal{V}$ and $\mathcal{V}'$ are the same.

The change of basis described in Eq.(8.5) of Section 8.1 leads us from one matrix representation to another equivalent one. In the definition of equivalence of representations given above, no statement about the reducibility or otherwise of $D(\cdot)$ and $D'(\cdot)$ was made. Actually one can easily convince oneself that two equivalent representations must be both irreducible, or both reducible; and in the latter case both must be decomposable or both indecomposable. For

irreducible representations, a very useful test of equivalence exists, and it is called Schur's Lemma. It states: if $D(\cdot)$ and $D'(\cdot)$ are irreducible representations of a group G on linear spaces $\mathcal{V}$ and $\mathcal{V}'$, and if there is a linear mapping $T : \mathcal{V} \to \mathcal{V}'$ that intertwines the two representations in the sense

$$D'(g)T = TD(g) \text{ for all } g \in G, \tag{8.13}$$

then either

(i) $D(\cdot)$ and $D'(\cdot)$ are inequivalent, and $T = 0$, or

(ii) $D(\cdot)$ and $D'(\cdot)$ are equivalent, and in case $T \neq 0$, then T^{-1} exists.

This is a very nice result, which we can exploit in more than one way. On the one hand, if we know that $D(\cdot)$ and $D'(\cdot)$ are inequivalent, then we are sure that any intertwining T we may succeed in constructing must be zero. On the other hand, if we know that T is non-zero, then in fact T will also be non-singular, and the representations must be equivalent. With both representations given to be irreducible, it can never happen that the intertwining operator T is non-zero and singular! The proof (which we do not give here) is a clever exploitation of the range and null spaces of T, respectively subspaces of $\mathcal{V}'$ and $\mathcal{V}$, and the assumed irreducibility of $D(\cdot)$ and $D'(\cdot)$.

As a special case of this general result, we can take $D'(\cdot) = D(\cdot)$, so we have just one irreducible representation on a given space $\mathcal{V}$. Then one easily finds the result:

$$\begin{aligned} &D(\cdot) \text{ irreducible, } D(g)T = TD(g) \text{ for all } g \in G \\ &\qquad \Rightarrow T = \lambda\, 1 = \text{ multiple of the unit operator on } \mathcal{V}. \end{aligned} \tag{8.14}$$

8.4 Unitary and Orthogonal Representations

We now consider spaces $\mathcal{V}$ carrying an inner product, and representations $D(\cdot)$ preserving this product. First we take the case of a complex linear space $\mathcal{V}$, later the real case.

Denote by (x, y) the hermitian nondegenerate positive definite inner product on the complex vector space $\mathcal{V}$. As is conventional in quantum mechanics, we assume it to be linear in y and antilinear in x. A representation $D(\cdot)$ of G on $\mathcal{V}$ being given, we say it is *unitary* if it respects the inner product, i.e.,

$$\begin{aligned} &(D(g)x, D(g)y) = (x, y) \text{ for all } g \in G; x, y \in \mathcal{V} \\ \text{i.e., } &(x, D(g)^\dagger D(g)y) = (x, y) \text{ for all } g \in G; x, y \in \mathcal{V} \\ \text{i.e., } &D(g)^\dagger D(g) = \mathbf{1} = \text{ unit operator on } \mathcal{V} \end{aligned} \tag{8.15}$$

The dagger denotes of course the hermitian adjoint determined by the inner product $(.\ ,.)$. If we use an orthonormal basis for $\mathcal{V}$, then the matrices $(D(g)_{jk})$ will be unitary matrices.

Next, consider a real vector space $\mathcal{V}$ carrying a symmetric nondegenerate positive definite inner product, which for simplicity is again denoted by (x, y). By introducing an orthonormal basis, and associating components with the vectors x and y, we can put the inner product into the usual form

$$(x, y) = x^T y \tag{8.16}$$

If the representation $D(\cdot)$ of G on $\mathcal{V}$ preserves this inner product, we say it is *real orthogonal*. The corresponding matrices then obey, in addition to being real,

$$D(g)^T D(g) = 1 \tag{8.17}$$

For finite groups, as well as for compact Lie groups, it is true that any representation is equivalent to a unitary one; we can say any representation is essentially unitary. If for a group, every representation is both essentially unitary and fully reducible, then the basic building blocks for its representation theory are unitary irreducible representations, often abbreviated as UIR's. This is again so for finite groups and for compact simple Lie groups.

Quite often, a complex vector space $\mathcal{V}$ and a representation $D(\cdot)$ of G on $\mathcal{V}$ may be given, and we may search for an acceptable inner product on $\mathcal{V}$ under which $D(\cdot)$ is unitary. If no such inner product can be found, then we have a *non-unitary* representation. Physically important examples are all the finite dimensional nontrivial representations of the Lorentz group SO(3, 1), of its covering group SL(2, C), and of the higher dimensional groups SO(p, q) of pseudo-orthogonal type.

It can very well happen that in a suitable basis for a complex vector space $\mathcal{V}$, the matrices $D(\cdot)$ of a representation all become real. This motivates questions like: when can a unitary representation be brought to real orthogonal form? We shall describe the answer after defining some common ways of passing from a given representation to related ones.

8.5 Contragredient, Adjoint and Complex Conjugate Representations

Let $D(\cdot)$ be an irreducible representation of G on $\mathcal{V}$; for definiteness choose some basis in $\mathcal{V}$, and work with the corresponding matrices of the representation. We can take $\mathcal{V}$ to be a complex space, but do not assume any inner product on it, or that $D(\cdot)$ is unitary. (The alternative to using a basis and representative matrices would be to work with the dual to $\mathcal{V}$). From this given irreducible representation, by simple matrix operations three other representations can be

constructed:

$$
\begin{aligned}
&g \to D(g) \text{ is an irreducible matrix representation} \Rightarrow \\
&g \to (D(g)^T)^{-1} \text{ is the irreducible representation } \textit{contragredient} \text{ to } D(\cdot); \\
&g \to (D(g)^\dagger)^{-1} \text{ is the irreducible representation } \textit{adjoint} \text{ to } D(\cdot); \\
&g \to D(g)^* \text{ is the irreducible representation } \textit{complex conjugate} \text{ to } D(\cdot)
\end{aligned} \tag{8.18}
$$

Since in this general situation, neither reality nor unitarity nor orthogonality of $D(\cdot)$ was assumed, each of these derived representations could be inequivalent to $D(\cdot)$ itself. If, however, $\mathcal{V}$ carried a hermitian inner product and $D(\cdot)$ was unitary, then not all these representations are different:

$$
\begin{aligned}
&D(\cdot) \text{ unitary } \Rightarrow D(\cdot) \text{ is } \textit{self-adjoint}, \\
&\textit{Contragredient} \equiv \textit{complex conjugate}
\end{aligned} \tag{8.19}
$$

Similarly, for a real space $\mathcal{V}, D(\cdot)$ is by definition real. In addition,

$$
D(\cdot) \text{ orthogonal } \Rightarrow D(\cdot) \text{ is } \textit{self-contragredient} \tag{8.20}
$$

In general terms, the property of being self-contragredient has interesting consequences. Take an irreducible matrix representation $D(\cdot)$ of G, whether real or complex, and suppose it is equivalent to its contragredient. This means that there is a non-singular matrix S such that

$$
(D(g)^T)^{-1} = SD(g)S^{-1} \text{ for all } g \in G \tag{8.21}
$$

By exploiting this relation twice, we quickly arrive at

$$
S^{-1}S^T D(g) = D(g)S^{-1}S^T \text{ for all } g \in G \tag{8.22}
$$

Then Schur's Lemma in the form (8.14) implies that for some constant c,

$$
\begin{aligned}
&S^{-1}S^T = c \cdot 1, \\
&i.e. \quad S^T = cS, \\
&i.e. \quad S = c^2 S, \\
&i.e. \quad c = \pm 1, S^T = \pm S
\end{aligned} \tag{8.23}
$$

Thus, the similarity transformation relating $D(\cdot)$ and $(D(\cdot)^T)^{-1}$ is definitely either symmetric or antisymmetric, and the latter case can only arise if the dimension of the representation is even. Let us now write the equivalence (8.21) in the form

$$
D(g)^T S D(g) = S \text{ for all } g \in G \tag{8.24}
$$

This means that if we define a non-degenerate bi-linear form (x, y) on $\mathcal{V}$ using the matrix S as

$$
(x, y) = x^T S y, \tag{8.25}
$$

then the representation $D(\cdot)$ preserves this form:

$$(D(g)x, D(g)y) = x^T D(g)^T S D(g) y = (x, y) \tag{8.26}$$

Thus, the characteristic of a self contragredient representation is that it preserves a symmetric or an antisymmetric nondegenerate bi-linear form (there is no statement however of positive definiteness of this form!). Conversely, invariance of such a bi-linear form evidently means that $D(\cdot)$ is self contragredient.

As a last point in this section, we analyse the possible ways in which an irreducible representation $D(\cdot)$ can be related to its complex conjugate $D(\cdot)^*$. We will deal only with the case where $D(\cdot)$ is given to be unitary as well. Even so, the situation shows some subtleties, and it is worth examining these in some detail.

We are given, then, a UIR $D(\cdot)$ of a group G on a complex linear vector space $\mathcal{V}$ equipped with a hermitian positive definite inner product. By choosing an orthonormal basis, we have unitary representation matrices $D(g)$. Let the dimension of $\mathcal{V}$ be n. There are three mutually exclusive possibilities for the relationship between $D(\cdot)$ and $D(\cdot)^*$:

(i) Complex case: $D(g)$ and $D(g)^*$ are inequivalent UIR's.

(ii) Potentially real case: $D(g)$ and $D(g)^*$ are equivalent, and one can choose an orthonormal basis for $\mathcal{V}$ such that $D(g)$ becomes real for all $g \in G$.

(iii) Pseudo real case: $D(g)$ and $D(g)^*$ are equivalent, but they cannot be brought to real form.

What we need to do is to characterise cases (ii) and (iii) in a useful and interesting way. Since we have assumed unitarity, in both cases we have a self-contragredient representation, so according to the discussion earlier in this section there is a symmetric or antisymmetric matrix S transforming $D^*(\cdot)$ into $D(\cdot)$:

$$\begin{aligned} D(g)^* &= S D(g) S^{-1} \\ S^T &= \pm S \end{aligned} \tag{8.27}$$

We will in fact find that

$$\begin{aligned} &\text{Case (ii), } D(g) \text{ potentially real} \Leftrightarrow S^T = S, \\ &\text{Case (iii), } D(g) \text{ pseudo real} \Leftrightarrow S^T = -S, \text{ so } n \text{ must be even} \end{aligned} \tag{8.28}$$

Case (iii) is simpler to handle than case (ii), so we dispose of it first. Maintaining the unitarity of the representation, the most general change we can make is a unitary transformation, which changes S in such a way as to maintain its antisymmetry (as we would have expected in any event):

$$\begin{aligned} D'(g) &= U D(g) U^{-1}, U^\dagger U = 1 \Rightarrow \\ D'(g)^* &= S' D'(g) (S')^{-1} \\ S' &= (U^T)^{-1} S U^{-1} \Rightarrow \\ (S')^T &= -S' \end{aligned} \tag{8.29}$$

Thus, no choice of U can replace S by the unit matrix or a multiple thereof, which by Schur's Lemma is what would have to be done if $D'(g)$ were to become explicitly real. We have therefore shown that

$$S^T = -S \Leftrightarrow D(\cdot) \text{ cannot be brought to real form} \tag{8.30}$$

Case (ii) is a little more intricate. Now, with no loss of generality, we have that S is both unitary and symmetric:

$$\begin{aligned} S^\dagger S = 1, S^T = S \Rightarrow \\ S^* = S^{-1} \end{aligned} \tag{8.31}$$

Our aim is to show that we can unitarily transform $D(\cdot)$ to $D'(\cdot)$ by a suitable choice of unitary U, as in the first line of Eq.(8.29), such that $D'(\cdot)$ is real. The key is to realise that because of the properties (8.31) of S, we can express S as the square of a unitary symmetric matrix U:

$$\begin{aligned} S = U^2 \\ U^\dagger U = 1, \qquad U^T = U \end{aligned} \tag{8.32}$$

The argument is the following: being unitary, the eigenvalues of S are all unimodular, and we can form an orthonormal basis out of the eigenvectors of S. If ψ is an eigenvector of S with eigenvalue $e^{i\varphi}$, then

$$\begin{aligned} S\psi = e^{i\varphi}\psi \Rightarrow S^*\psi^* = e^{-i\varphi}\psi^* \\ \Rightarrow S\psi^* = e^{i\varphi}\psi^* \end{aligned} \tag{8.33}$$

on account of Eq.(8.31). Therefore either ψ^* is proportional to ψ, in which case its phase can be adjusted to make ψ real; or else the real and imaginary parts of ψ are both eigenvectors of S with the same eigenvalue $e^{i\varphi}$. In any case, we see that the eigenvectors of S can all be chosen real, and this is also true for the members of an orthonormal basis of eigenvectors. This means that S can be diagonalised by a real orthogonal matrix O:

$$\begin{aligned} S = OdO^{-1}, \\ O^TO = 1, \quad O^* = O, \\ d = \text{ diagonal, unitary} \end{aligned} \tag{8.34}$$

If we now take $d^{1/2}$ to be any square root of d, likewise diagonal and unitary, we see that

$$U = Od^{1/2}O^{-1} \tag{8.35}$$

obeys all of Eqs.(8.32). If the expression U^2 is used for S in Eq.(8.27), after rearrangement of factors, we get

$$(UD(g)U^{-1})^* = UD(g)U^{-1} \tag{8.36}$$

showing that the representation $D(\cdot)$ has been brought to real form.

The most familiar illustration of all this is in the representation theory of SU(2), or angular momentum theory in quantum mechanics. It is well-known that all integer spin representations, which are odd-dimensional, are potentially real, and indeed their description using Cartesian tensors gives them in real form. However, all half odd integer spin representations which are even dimensional, are pseudo real: the invariant bi-linear form in these cases is antisymmetric.

8.6 Direct Products of Group Representations

As a last item in this brief review of group representations, let us look at the process of forming the direct product of two irreducible representations of a group G. Let the two representations $D_1(\cdot)$ and $D_2(\cdot)$ operate on vector spaces $\mathcal{V}_1$ and $\mathcal{V}_2$ respectively. Assume they are both complex or both real. First set up the direct product vector space $\mathcal{V}_1 \times \mathcal{V}_2$: it is again a complex or a real space depending on the natures of $\mathcal{V}_1$ and $\mathcal{V}_2$. As is well-known, $\mathcal{V}_1 \times \mathcal{V}_2$ is spanned by outer products or ordered pairs of vectors of the form $x \otimes y$, where $x \in \mathcal{V}_1$, and $y \in \mathcal{V}_2$. (The building up of $\mathcal{V}_1 \times \mathcal{V}_2$ is rather subtle, if one wants to do it in an intrinsic manner, but we do not go into those details here!). Then we define the direct product representation $D(\cdot) = D_1(\cdot) \times D_2(\cdot)$ by giving the action on a "monomial" as

$$D(g)(x \otimes y) = D_1(g)x \otimes D_2(g)y \tag{8.37}$$

and then extending the action to all of $\mathcal{V}_1 \times \mathcal{V}_2$ by linearity. Evidently, we do have a representation of G here.

Even if $D_1(\cdot)$ and $D_2(\cdot)$ are irreducible, $D(\cdot)$ may be reducible. If it is fully reducible, in its reduction various irreducible representations may occur with various multiplicities. This is the *Clebsch-Gordan problem*, and the direct sum of irreducible representations contained in the product $D_1(\cdot) \times D_2(\cdot)$ is called the *Clebsch-Gordan series*.

Exercises for Chapter 8

1. For the n-element cyclic group $C_n = \{e, a, a^2, \ldots, a^{n-1}; a^n = e\}$, find all irreducible unitary (one-dimensional) representations.

2. For the group of real translations $x \to x + a$ in one dimension, show that

$$a \longrightarrow \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$$

is a representation. Is it reducible or irreducible, decomposable or indecomposable?

3. Find all one dimensional unitary (irreducible) representations of the n dimensional real translation group $x_j \to x_j + a_j, j = 1, 2, \ldots, n$.

4. Show that the group in problem (3) is represented unitarily on the space of all square integrable complex functions $f(\underline{x})$ on R^n by

$$f(\underline{x}) \to (T(a)f)(\underline{x}) = f(\underline{x} - \underline{a}).$$

Chapter 9

Lie Groups and Lie Algebras

We shall in this Chapter study the properties of Lie groups in some detail, using local systems of coordinates rather than the admittedly more powerful and concise intrinsic geometric methods. However as an exercise the reader is invited to apply the concepts of differential geometry developed elsewhere in this monograph, to see for himself or herself how to make this local description a global one. We shall see how Lie algebras arise from Lie groups, and how one can make the reverse transition as well.

9.1 Local Coordinates in a Lie Group

Let there be a Lie group G of dimension r. Then in some suitable neighbourhood $\mathfrak{N}$ of the identity e, we are able to smoothly assign coordinates in Euclidean r-dimensional space to group elements. The dimension being r implies that these coordinates are both *essential* (i.e., we cannot do with less) and *real*. We use Greek letters for coordinates and use the convention,

$$\begin{aligned} a \in \mathfrak{N} \subset G: &\quad \text{coordinates } \alpha^1, \alpha^2, \ldots, \alpha^r \equiv \alpha^j, j = 1, 2, \ldots, r; \\ b \in \mathfrak{N} \subset G: &\quad \text{coordinates } \beta^1, \beta^2, \ldots, \beta^r \equiv \beta^j, j = 1, 2, \ldots, r \end{aligned} \tag{9.1}$$

As a convention we always agree to assign coordinates zero to the identity:

$$a = e : \alpha^j = 0 \tag{9.2}$$

As the variable element a runs over the open set $\mathfrak{N}$, α^j runs over some open set around the origin in Euclidean r-space. In this region, group elements and coordinates *determine each other uniquely.*

The freedom to make smooth changes of coordinates in an invertible way, $\alpha^j \to \tilde{\alpha}^j$ say, is of course available. We must only observe that

$$\alpha^j = 0 \Leftrightarrow \tilde{\alpha}^j = 0 \tag{9.3}$$

Provided the elements $a, b \in \mathfrak{N}$ are chosen so that the product $c = ab \in \mathfrak{N}$ as well (and similarly later on if we need to compose more elements), the law of group composition in G is given by a set of r real-valued functions of $2r$ real arguments each:

$$c = ab \Leftrightarrow \gamma^j = f^j(\alpha; \beta), \quad j = 1, 2, \ldots, r \tag{9.4}$$

Thus in each system of coordinates, the structure of G is expressed in a corresponding system of *group composition functions*. Our convention (9.2) means that these functions obey

$$f^j(\alpha; 0) = f^j(0; \alpha) = \alpha^j \tag{9.5}$$

The existence of a unique inverse, a^{-1}, to each $a \in \mathfrak{N}$ (provided a is chosen so that $a^{-1} \in \mathfrak{N}$ as well!) is expressed as the following property of the f's: Given α^j, there is a unique α'^j such that

$$f^j(\alpha; \alpha') = f^j(\alpha'; \alpha) = 0, a \to \alpha, a^{-1} \to \alpha' \tag{9.6}$$

A change of coordinates (9.3) will result in a new set of functions $\tilde{f}^j(.;.)$, say, still subject to Eqs (9.5), (9.6). We will assume for definiteness that the f's are sufficiently often differentiable; (that this need not be assumed is an important result in the theory of Lie groups, but we are working at a less sophisticated level !). Thus smooth coordinate changes will guarantee this property for $\tilde{f}$ as well.

We now study the properties of f^j implied by the fact that G is a group, in a step-by-step manner.

9.2 Analysis of Associativity

Take three group elements $a, b, c \in \mathfrak{N}$ such that the product $cab \in \mathfrak{N}$ as well. Associativity leads to functional equations for the f's:

$$\begin{gathered} c(ab) = (ca)b \Rightarrow \\ f^j(\gamma; f(\alpha; \beta)) = f^j(f(\gamma; \alpha); \beta) \end{gathered} \tag{9.7}$$

Differentiate both sides with respect to γ^k and then set $\gamma = 0$, i.e. $c = e$:

$$\frac{\partial f^j}{\partial \gamma^k}(\gamma; f(\alpha; \beta))|_{\gamma=0} = \frac{\partial f^j}{\partial \alpha^l}(\alpha; \beta) \frac{\partial f^l}{\partial \gamma^k}(\gamma; \alpha)|_{\gamma=0} \tag{9.8}$$

Here we have used Eq.(9.5) for the first factor on the right. Now the expressions on the left and on the extreme right motivate us to define a system of r^2 functions of r real arguments each in this way:

$$\eta_k^j(\alpha) = \frac{\partial f^j}{\partial \gamma^k}(\gamma;\alpha)|_{\gamma=0} \tag{9.9}$$

Thus, while the f's are functions of two independent group elements, the η's have only one group element as argument. Again on account of Eq.(9.5), we have

$$\eta_k^j(0) = \delta_k^j \tag{9.10}$$

so we assume $\mathfrak{N}$ is so chosen as to make $\eta_k^j(\alpha)$ as a matrix nonsingular all over $\mathfrak{N}$:

$$\begin{aligned} H(\alpha) &= (\eta_k^j(\alpha)) \\ H(\alpha)^{-1} &= \Xi(\alpha) = (\xi_k^j(\alpha)) \\ \eta_k^j(\alpha)\xi_l^k(\alpha) &= \delta_l^j, \\ \xi_k^j(0) &= \delta_k^j \end{aligned} \tag{9.11}$$

We treat superscripts (subscripts) as row (column) indices for matrix operations.

We can say: if a system of group composition functions obeying the associativity law is given, then matrices $H(\alpha), \Xi(\alpha)$ inverse to one another can be defined; and the f's will then obey, as a consequence of associativity, the (generally nonlinear) system of partial differential equations

$$\begin{aligned} \frac{\partial f^j}{\partial \alpha^k}(\alpha;\beta) &= \eta_l^j(f(\alpha;\beta))\xi_k^l(\alpha) \\ &= (H(f(\alpha;\beta))\Xi(\alpha))_k^j, \\ f^j(0;\beta) &= \beta^j \end{aligned} \tag{9.12}$$

Here the α's are the active independent variables while the β's appear via boundary conditions.

We can now exploit this system (9.12) to our advantage. Suppose the composition functions f^j are not known, but instead the functions η, ξ are given, and one is also assured that the partial differential equations (9.12) can be solved for the f's. Then the structure of these equations guarantees that, once the f's have been found, they will have the associativity property, will provide inverses etc, so that they will describe a group!

It is perhaps useful to describe this situation in another way. If some "arbitrary" nonsingular matrix of functions (η) is given to us with inverse matrix (ξ), and we then set up the system (9.12) for unknown f's, these equations will in general not be soluble at all ! The η's must possess definite properties if eqs (9.12) are to possess a solution; these will be obtained in Section 9.4. But assuming for now that an *acceptable* set of η's are given, we can show that the

f's do have the correct properties to describe a group (at least locally!). This we now prove: basically one just exploits the fact that (η) and (ξ) are inverses to one another.

Write $L^j(\gamma)$ and $R^j(\gamma)$ for the expressions on the two sides of Eq.(9.7), and regard γ as the active variables while both α and β are left implicit. We can develop systems of partial differential equations and boundary conditions for both $L(\gamma)$ and $R(\gamma)$, based on Eqs (9.11),(9.12):

$$\begin{aligned}
L^j(\gamma) &\equiv f^j(\gamma; f(\alpha;\beta)) : \\
\frac{\partial L^j}{\partial \gamma^k}(\gamma) &= (H(L)\Xi(\gamma))^j_k, \\
L^j(0) &= f^j(\alpha;\beta); \\
R^j(\gamma) &\equiv f^j(f(\gamma;\alpha);\beta) : \\
\frac{\partial R^j}{\partial \gamma^k}(\gamma) &= \frac{\partial f^j}{\partial \delta^l}(\delta;\beta)|_{\delta=f(\gamma;\alpha)} \frac{\partial f^l}{\partial \gamma^k}(\gamma;\alpha) \\
&= (H(R)\Xi(f(\gamma;\alpha))H(f(\gamma;\alpha))\Xi(\gamma))^j_k, \\
&= (H(R)\Xi(\gamma))^j_k, \\
R^j(0) &= f^j(\alpha;\beta)
\end{aligned} \tag{9.13}$$

Since $L(\gamma)$ and $R(\gamma)$ obey the same partial differential equations and boundary conditions with respect to γ, and these equations have a unique solution, we must have

$$L^j(\gamma) = R^j(\gamma) \tag{9.14}$$

Thus associativity is guaranteed.

How about inverses for group elements? Since (η) and (ξ) are both nonsingular, so is the Jacobian matrix $(\partial f(\alpha;\beta)/\partial\alpha)$. So given α, there is a unique α' such that

$$f^j(\alpha;\alpha') = 0 \tag{9.15}$$

(Remember again that here we are viewing the η's as given, have found the f's by solving Eqs.(9.12), and are establishing various properties for them!). These α' are the coordinates of a^{-1}, where α are the coordinates of a. But while we have determined α' by imposing the condition $aa^{-1} = e$, will we also have $a^{-1}a = e$? In other words, will the α' obtained by solving Eq.(9.15) also obey $f(\alpha';\alpha) = 0$? Yes, because of the associativity law (9.14) already proved! Using the rule (9.15) twice, let us suppose we have:

$$\begin{aligned}
a \sim \alpha \to f^j(\alpha;\alpha') = 0 \to a^{-1} \sim \alpha'; \\
a^{-1} \sim \alpha' \to f^j(\alpha';\alpha'') = 0 \to (a^{-1})^{-1} \sim \alpha''
\end{aligned} \tag{9.16}$$

Then associativity gives:

$$\begin{aligned} f(\alpha; f(\alpha'; \alpha'')) &= f(f(\alpha; \alpha'); \alpha''), \\ \text{i.e., } \alpha &= \alpha'', \\ \text{i.e., } a &= (a^{-1})^{-1}, \\ \text{i.e., } a^{-1}a &= a^{-1}(a^{-1})^{-1} = e. \end{aligned} \tag{9.17}$$

To sum up, if a Lie group G and a coordinate system α are given, then the f's are known, the η's and ξ's can be computed, and they will of course be *acceptable*. Conversely, if acceptable η's and ξ's are given, we can build up the f's uniquely, and thus reconstruct the group, at least locally.

The importance of the systems of functions η, ξ must be evident. An interpretation for the η's can be developed thus:

Following from the definition (9.9), for small α we can write

$$f^j(\alpha; \beta) \simeq \beta^j + \eta_k^j(\beta)\alpha^k + 0(\alpha^2) \tag{9.18}$$

So, if a is an element close to the identity e, with coordinates α, and b is some (finite) element with coordinates β, then ab is close to b and has coordinates which differ from β by the amount $H(\beta)\alpha$:

$$\begin{aligned} a \text{ near } e, b \text{ general } &\Rightarrow b' = ab \text{ near } b, \\ &\beta' \simeq \beta + H(\beta)\alpha + 0(\alpha^2) \end{aligned} \tag{9.19}$$

9.3 One-parameter Subgroups and Canonical Coordinates

A one parameter subgroup (OPS) in a Lie group G is a continuous and differentiable family of elements $a(\sigma) \in G$, where σ is a real parameter, forming a subgroup:

$$\begin{aligned} a(\sigma)a(\sigma') &= a(\sigma + \sigma'), \\ a(0) &= e, \\ a(\sigma)^{-1} &= a(-\sigma) \end{aligned} \tag{9.20}$$

(It can be shown that an OPS, defined more weakly than here, is necessarily abelian, and that the parameter can be chosen so that group composition corresponds to parameter addition; for simplicity, all this has been made part of the definition of an OPS).

We now subject this concept to the same kind of treatment as the group itself in the previous section. At first, given G, and an OPS in it, we develop a way to *describe* the latter. Then later we turn the situation around, and use the *description* as a way to *determine* an OPS (G being given).

Given G and $a(\sigma)$, in the coordinate system α let $a(\sigma)$ have coordinates $\alpha(\sigma)$:

$$\begin{aligned} a(\sigma) &\to \alpha^j(\sigma); \\ \alpha^j(0) &= 0; \\ f(\alpha(\sigma'); \alpha(\sigma)) &= \alpha(\sigma' + \sigma) \end{aligned} \tag{9.21}$$

Now take σ' to be an infinitesimal $\delta\sigma$, and also write

$$\begin{aligned} t^j &= \frac{d\alpha^j}{d\sigma}(\sigma)|_{\sigma=0}, \\ \alpha^j(\delta\sigma) &\simeq t^j \delta\sigma \end{aligned} \tag{9.22}$$

Then from Eq.(9.21),

$$\begin{aligned} \alpha^j(\sigma + \delta\sigma) &= f^j(\alpha(\delta\sigma); \alpha(\sigma)) \\ &\simeq f^j(t\delta\sigma; \alpha(\sigma)) \\ &\simeq \alpha^j(\sigma) + \eta_k^j(\alpha(\sigma)) t^k \delta\sigma \end{aligned} \tag{9.23}$$

so the coordinates of the elements on an OPS obey the (generally nonlinear) ordinary differential equations and boundary conditions

$$\begin{aligned} \frac{d\alpha^j}{d\sigma}(\sigma) &= \eta_k^j(\alpha(\sigma)) t^k, \\ \alpha^j(0) &= 0 \end{aligned} \tag{9.24}$$

We say t^j are the components of the tangent vector to the OPS at the identity.

Can the above be reversed? Given G, the f's, η's and ξ's, suppose *we choose* an r-component real vector t^j, and solve the system (9.24) and get functions $\alpha^j(\sigma)$. Will these be the coordinates on an OPS and will its tangent at the identity be t? The answer is in the affirmative, and the proof is similar to that in the previous section. We wish to prove that, given Eqs.(9.24) with solution $\alpha^j(\sigma)$,

$$L^j(\sigma) \equiv \alpha^j(\sigma + \sigma') = f^j(\alpha(\sigma); \alpha(\sigma')) \equiv R^j(\sigma) \tag{9.25}$$

Here we regard σ as active and σ' as implicit in L and R. Then eqs (9.24),(9.12) give:

$$\begin{aligned} \frac{dL^j}{d\sigma}(\sigma) &= \eta_k^j(L(\sigma)) t^k, \\ L^j(0) &= \alpha^j(\sigma'); \\ \frac{dR^j}{d\sigma}(\sigma) &= (H(R(\sigma)) \Xi(\alpha(\sigma)) H(\alpha(\sigma)))_k^j t^k, \\ &= \eta_k^j(R(\sigma)) t^k, \\ R^j(0) &= \alpha^j(\sigma') \end{aligned} \tag{9.26}$$

Since both the differential equations and boundary conditions are the same for L and R, they coincide and the equality (9.25) is proved.

In summary, a given OPS determines a tangent vector t uniquely and vice versa, based on the differential equations (9.24).

So far we have allowed considerable flexibility in choice of coordinate systems α for G. The properties of OPS's motivate the definition of a special family of coordinate systems, the so called canonical coordinates (of the first kind). At first, let us denote the solution to Eqs.(9.24) by $\alpha(\sigma;t)$, indicating explicitly the dependence on the tangent vector t: there are $r+1$ real arguments here. But a moment's reflection shows that the dependence can only be on the r products $\sigma t^1, \sigma t^2, \ldots, \sigma t^r$, so we will write the solution to Eqs.(9.24) as:

$$\begin{aligned} \alpha^j(\sigma;t) &\equiv \chi^j(\sigma t) \\ &\equiv r \text{ functions of } r \text{ real independent arguments;} \\ \chi^j(0) &= 0, \\ \chi^j(t\delta\sigma) &\simeq t^j\delta\sigma + 0(\delta\sigma)^2. \end{aligned} \tag{9.27}$$

In fact the differential equations (9.24) for an OPS can be expressed as a system of equations obeyed by the functions χ:

$$\begin{aligned} t^k\frac{\partial\chi^j}{\partial t^k}(t) &= \eta^j_k(\chi(t))t^k \\ \chi(0) &= 0 \\ \frac{\partial\chi^j}{\partial t^k}(t)|_{t=0} &= \delta^j_k \end{aligned} \tag{9.28}$$

Thus, given G and a coordinate system α around e, we have arrived at the system of functions $\chi(t)$. Now the Jacobian property of χ above means that if any element $a \in \mathfrak{N}$ (sufficiently small) with coordinates α^j is given, we can find a unique vector t^j such that

$$\chi^j(t) = \alpha^j \tag{9.29}$$

In other words: for any $a \in \mathfrak{N}$, there is a *unique* tangent vector t such that the corresponding OPS passes through a at parameter value $\sigma = 1$. We express this relationship in the (at this point somewhat formal) *exponential notation*:

a with coordinates α^j : $a = \exp(t)$,

$$\alpha^j = \alpha^j(\sigma;t)|_{\sigma=1} = \chi^j(t). \tag{9.30}$$

Thus in this notation the elements $a(\sigma)$ on an OPS with tangent vector t are $\exp(\sigma t)$, and the OPS property means

$$\exp(\sigma t)\exp(\sigma' t) = \exp((\sigma+\sigma')t). \tag{9.31}$$

All $a \in \mathfrak{N}$ are obtained by suitably varying t in $\exp(t)$. In particular,

$$\begin{aligned} e &= \exp(0), \\ (\exp(t))^{-1} &= \exp(-t) \end{aligned} \tag{9.32}$$

Can we now exploit these results to pass from the initial arbitrarily chosen coordinates α^j to a *new* system of coordinates in which not only are t^j the components of the tangent vector to the OPS passing through a given $a \in \mathfrak{N}$ at $\sigma = 1$, but also serve as the coordinates of a? Yes, because once the functions $\chi(t)$ are in hand, we can in principle turn the equations

$$\alpha^j = \chi^j(t) \tag{9.33}$$

"inside out" (the Jacobian property of χ in Eq. (9.28) ensures this), and express t in terms of α:

$$t^j = \zeta^j(\alpha) \equiv \text{ new coordinates } \bar{\alpha}^j \tag{9.34}$$

The distinguishing characteristics of these new coordinates $\bar{\alpha}$ for $\mathfrak{N} \subset G$ are that on an OPS $a(\sigma)$:

$$a(\sigma) \rightarrow \text{ old coordinates } \alpha^j(\sigma) = \chi^j(\sigma t) \rightarrow$$

$$\text{new coordinates } \bar{\alpha}^j(\sigma) = \sigma t^j \tag{9.35}$$

Such coordinates are called *canonical coordinates*. If we were to write $\bar{f}, \bar{\eta}, \bar{\xi}$ for the functions associated with them, the characterisation (9.35) means that now the solution to Eqs.(9.24) in the new coordinates is very simple:

$$\begin{aligned} &\frac{d\bar{\alpha}^j}{d\sigma}(\sigma) = \bar{\eta}^j_k(\bar{\alpha}(\sigma))t^k, \bar{\alpha}^j(0) = 0 \Rightarrow \\ &\bar{\alpha}^j(\sigma) = \sigma t^j \Rightarrow \\ &\bar{\eta}^j_k(\bar{\alpha})\bar{\alpha}^k = \bar{\alpha}^j, \\ \text{i.e.,} \quad &\bar{H}(\bar{\alpha})\bar{\alpha} = \bar{\alpha}. \end{aligned} \tag{9.36}$$

We thus have two direct ways of defining a canonical coordinate system: either (i) the functions $\bar{\eta}$ must have the functional property appearing in Eq.(9.36) or (ii) the elements on an OPS have coordinates σt.

The passage from a general coordinate system α to the canonical system $\bar{\alpha}$ *uniquely associated with it* has the feature that upto linear terms they agree, and (possibly) diverge only thereafter:

$$\begin{aligned} \bar{\alpha}^j &\simeq \alpha^j + 0(\alpha^2), \\ \alpha^j &\simeq \bar{\alpha}^j + 0(\bar{\alpha}^2) \end{aligned} \tag{9.37}$$

(In particular, if α were already canonical, then $\bar{\alpha} = \alpha$!) We can say that while there are "infinitely many" general coordinate systems for $\mathfrak{N} \subset G$ related to one another by *functional transformations*, there is a many-to-one passage from general to canonical coordinate systems; these are "far fewer" but still infinite in number, and related to one another by *linear transformations*. Of course the set of canonical coordinate systems appears already as a subset of the set of all general coordinate systems; and all these relationships hinge on the concept and properties of OPS's.

9.4 Integrability Conditions and Structure Constants

Now we consider the conditions on the η's and ξ's which ensure that Eqs.(9.12) for f's can be solved. Regarding the β's as fixed parameters, these integrability conditions are:

$$\frac{\partial}{\partial\alpha^m}\frac{\partial}{\partial\alpha^k}f^j = \frac{\partial}{\partial\alpha^k}\frac{\partial}{\partial\alpha^m}f^j,$$

$$\text{i.e.} \qquad \frac{\partial}{\partial\alpha^m}(\eta_l^j(f)\xi_k^l(\alpha)) = \frac{\partial}{\partial\alpha^k}(\eta_l^j(f)\xi_m^l(\alpha)) \tag{9.38}$$

We can develop both sides using Eqs.(9.12) again, and then rearrange terms so that all expressions evaluated at f are on one side, those evaluated at α on the other. Partial derivatives are indicated by indices after a comma, and the result is:

$$\begin{aligned}&\xi_m^l(f)[\eta_j^n(f)\eta_{k,n}^m(f) - \eta_k^n(f)\eta_{j,n}^m(f)]\\ &\quad = \eta_j^m(\alpha)\eta_k^n(\alpha)[\xi_{m,n}^l(\alpha) - \xi_{n,m}^l(\alpha)]\end{aligned} \tag{9.39}$$

Since the two sides are evaluated at different points in $\mathfrak{N}$ (remember the β's!), *each side must be a constant indexed by j, k and l.* Let us write $-c_{jk}^l$ for these constants: then the integrability conditions for Eqs.(9.12) are that there must be these constants and

$$\begin{aligned}\xi_m^l(\alpha)[\eta_j^n(\alpha)\eta_{k,n}^m(\alpha) - \eta_k^n(\alpha)\eta_{j,n}^m(\alpha)] &= -c_{jk}^l, \qquad \text{(a)}\\ \eta_j^m(\alpha)\eta_k^n(\alpha)[\xi_{m,n}^l(\alpha) - \xi_{n,m}^l(\alpha)] &= -c_{jk}^l \qquad \text{(b)}\end{aligned} \tag{9.40}$$

These two conditions are identical in content, on account of $(\eta)^{-1}$ being (ξ). It is convenient to express them in the forms

$$\begin{aligned}&\eta_j^n(\alpha)\eta_{k,n}^m(\alpha) - \eta_k^n(\alpha)\eta_{j,n}^m(\alpha) = -c_{jk}^l\eta_l^m(\alpha) \qquad \text{(a)}\\ &\Delta_{mn}^l(\alpha) \equiv \xi_{m,n}^l(\alpha) - \xi_{n,m}^l(\alpha) + c_{jk}^l\xi_m^j(\alpha)\xi_n^k(\alpha) = 0 \qquad \text{(b)}\end{aligned} \tag{9.41}$$

An "acceptable" set of functions $\eta(\alpha)$ must obey these partial differential equations for some numerical constants c_{jk}^l; if they do, then we can solve Eqs.(9.12) to get $f^j(\alpha;\beta)$, and these will determine the group structure (at least locally).

The c_{jk}^l are called "structure constants" for the group G. What are *acceptable* sets of structure constants? They *cannot* be chosen freely. If the group G were given, the corresponding c_{jk}^l can be calculated, and will be found (as we shall see) to obey certain algebraic conditions. In section 9.6, we will see that all sets of constants obeying these algebraic conditions *are* acceptable.

We outline in sequence the properties of the r^3 constants c_{jk}^l, which by definition are real numbers, assuming that G is given:

(i) Antisymmetry in the subscripts: obvious from Eqs.(9.40):

$$c^l_{jk} = -c^l_{kj} \tag{9.42}$$

(ii) Taking $\alpha = 0$ in Eq.(9.41a), we have the explicit expression

$$\begin{aligned} c^l_{jk} &= \eta^l_{j,k}(0) - \eta^l_{k,j}(0) \\ &= \left(\frac{\partial}{\partial \beta^j} \frac{\partial}{\partial \alpha^k} - \frac{\partial}{\partial \beta^k} \frac{\partial}{\partial \alpha^j} \right) f^l(\beta;\alpha)|_{\alpha=\beta=0} \end{aligned} \tag{9.43}$$

Therefore if we expand $f(\beta;\alpha)$ in a Taylor series as

$$f^l(\beta;\alpha) = \beta^l + \alpha^l + a^l_{jk}\beta^j\alpha^k + \cdots, \tag{9.44}$$

then,

$$c^l_{kj} = a^l_{kj} - a^l_{jk} \tag{9.45}$$

(iii) Relation to commutators of group elements: take two elements $a, b \in \mathfrak{N}$ both close to e, and calculate the coordinates of the commutator $q(a,b)$ (Eq.(7.16)): retaining leading terms only, we find:

$$\begin{aligned} &c = q(a,b): \\ &\gamma^j = c^j_{kl}\alpha^k\beta^l + \ldots \end{aligned} \tag{9.46}$$

This shows that if G is abelian, all the structure constants vanish; the converse is also true, as we will find later.

(iv) *Jacobi condition*: we exploit the integrability condition (9.41a) on η by multiplying it by two tangent vectors $t^j u^k$ and writing terms in a suggestive form:

$$t^j\eta^n_j(\alpha)\frac{\partial}{\partial\alpha^n}u^k\eta^m_k(\alpha) - u^k\eta^n_k(\alpha)\frac{\partial}{\partial\alpha^n}t^j\eta^m_j(\alpha) = -c^l_{jk}t^ju^k\eta^m_l(\alpha) \tag{9.47}$$

From experience in handling linear first order differential operators in quantum mechanics, we see that we can express this as a *commutation relation*:

$$\left[t^j\eta^n_j(\alpha)\frac{\partial}{\partial\alpha^n}, u^k\eta^m_k(\alpha)\frac{\partial}{\partial\alpha^m}\right] = -v^l\eta^m_l(\alpha)\frac{\partial}{\partial\alpha^m}, \quad v^l \equiv c^l_{jk}t^ju^k \tag{9.48}$$

That is, if for each tangent vector t we define the linear operator

$$\eta_{[t]} = t^j\eta^m_j(\alpha)\frac{\partial}{\partial\alpha^m}, \tag{9.49}$$

then the integrability conditions on η are,

$$\begin{aligned} [\eta_{[t]}, \eta_{[u]}] &\equiv \eta_{[t]}\eta_{[u]} - \eta_{[u]}\eta_{[t]} \\ &= \eta_{[v]} \end{aligned} \tag{9.50}$$

where v is another tangent vector formed out of t and u using the structure constants:

$$\begin{aligned} v &= [t,u] = \text{Lie bracket of } t \text{ with } u, \\ v^l &= c^l_{jk} t^j u^k \end{aligned} \tag{9.51}$$

We have introduced here the notion of the *Lie bracket*, an algebraic operation determining a tangent vector $[t,u]$ once t and u are given, at this moment to be distinguished from the commutator bracket occurring in Eq.(9.50). But, since commutators among operators do obey the Jacobi identity, Eqs.(9.50),(9.51) lead to the result that Lie brackets must also obey them! For any three vectors t, u and w, we have

$$\begin{gathered} \sum_{\text{cyclic}} [[\eta_{[t]}, \eta_{[u]}], \eta_{[w]}] = 0 \Rightarrow \\ \sum_{\text{cyclic}} [\eta_{[t,u]}, \eta_{[w]}] = 0 \Rightarrow \\ \sum_{\text{cyclic}} \eta_{[[t,u],w]} = 0 \Rightarrow \\ [[t,u],w] + [[u,w],t] + [[w,t],u] = 0 \end{gathered} \tag{9.52}$$

Here we have used the fact that $\eta_{[t]}$ depends linearly on t. Directly in terms of the structure constants, the Jacobi condition is

$$c^j_{kl} c^l_{mn} + c^j_{ml} c^l_{nk} + c^j_{nl} c^l_{km} = 0, \text{ all } jkmn \tag{9.53}$$

The complete set of conditions on structure constants which makes them acceptable are: reality, antisymmetry, Eq.(9.42); and the Jacobi identities (9.53). *There are no more conditions.* All these properties of structure constants, reflected also in the way the tangent vector v is formed out of t and u, motivate us to define the concept of the Lie Algebra.

9.5 Definition of a (real) Lie Algebra: Lie Algebra of a given Lie Group

A (real) Lie algebra is a real linear vector space, $\mathcal{L}$ say, with elements $u, v, t, w, \ldots$ among which a *Lie bracket* is defined. This is a rule which associates with any two given elements of $\mathcal{L}$ a third element, subject to certain conditions:

$$\begin{aligned} &\text{(i)} && u, v \in \mathcal{L} \Rightarrow [u,v] \in \mathcal{L}; \\ &\text{(ii)} && [u,v] = -[v,u]; \\ &\text{(iii)} && [u,v] \text{ linear in both } u \text{ and } v; \\ &\text{(iv)} && u,v,w \in \mathcal{L} \Rightarrow [[u,v],w] + [[v,w],u] + [[w,u],v] = 0 \end{aligned} \tag{9.54}$$

If a basis $\{e_j\}, j = 1, 2, \ldots, r$ is chosen for $\mathcal{L}$, the dimension of $\mathcal{L}$ being r, then because of the linearity property (9.54) (iii),

$$\begin{aligned} u = u^j e_j, v = v^j e_j \Rightarrow \\ [u, v] &= u^j v^k [e_j, e_k] \\ &= u^j v^k c^l_{jk} e_l, \\ [e_j, e_k] &= c^l_{jk} e_l, \\ \text{i.e.,} \qquad [u, v]^l &= c^l_{jk} u^j v^k \end{aligned} \tag{9.55}$$

For these structure constants c^l_{jk} of $\mathcal{L}$, properties (9.54) (ii),(iv)) then immediately give back the properties (9.42),(9.53) we had found for the structure constants of a Lie group!

We thus see that the definition of a Lie algebra is designed to capture in intrinsic form all those properties that are known to be possessed by the structure constants of any Lie group, in particular the process by which in Eqs.(9.50),(9.51), any two tangent vectors determine a third. We can now say: given a Lie group G of dimension r, the tangent vectors $u, v, w, t, \ldots$ at e to all possible OPS's in G taken together form an r-dimensional Lie algebra determined by G. We call it the *Lie algebra of* G. It is sometimes denoted by $\underline{G}$, sometimes by $\underline{g}$. Changing the coordinate system α over $\mathfrak{N} \subset G$ for the calculation of the structure constants c^l_{jk} results in a linear transformation, or rather a change of basis, in $\underline{G}$.

It is the group structure for elements in G infinitesimally close to e that determines the bracket structure in $\underline{G}$. Therefore if we have two (or more) Lie groups $G, G', \ldots$ which are *locally isomorphic* to one another in suitable neighbourhoods of their respective identity elements, they will all lead to, and so they will all share, the *same* Lie algebra $\mathcal{L}$. One can thus at best expect that a given Lie algebra can determine only the local structure of any of the Lie groups that could have led to it. We discuss this reconstruction process in the next section.

9.6 Local Reconstruction of Lie Group from Lie Algebra

Suppose a Lie algebra $\mathcal{L}$, that is, an acceptable set of structure constants c^l_{jk}, is given. Can we then say that we can compute the functions $\eta^j_k(\alpha)$? If the answer were in the affirmative, then the analysis in Section 9.2 tells us that we can go on to compute $f(\alpha; \beta)$ too, so from the structure constants we would have built up a Lie group G at least locally.

But actually *we know in advance* that the $\eta^j_k(\alpha)$ cannot be found if only c^l_{jk} are given! Suppose you already had a Lie group G in hand; clearly there is infinite freedom to change the coordinate system on $\mathfrak{N} \subset G$ leaving the coordinates around e unchanged to first order only. Such changes of coordinates

over $\mathfrak{N}$ would definitely change the $\eta_k^j(\alpha)$ but not the c_{jk}^l. So unless we restrict the coordinate system in some way, we cannot hope to get $\eta_k^j(\alpha)$ starting from c_{jk}^l.

The key is to use canonical coordinates of the first kind. It is possible, given an acceptable set of structure constants c_{jk}^l, to solve uniquely the combined system of partial differential equations and algebraic equations,

$$\begin{aligned} &\Delta_{mn}^l(\alpha) \equiv \xi_{m,n}^l(\alpha) - \xi_{n,m}^l(\alpha) + c_{jk}^l \xi_m^j(\alpha)\xi_n^k(\alpha) = 0, \\ &\xi_k^j(\alpha)\alpha^k = \alpha^j, \\ &\quad \xi_k^j(0) = \delta_k^j \end{aligned} \tag{9.56}$$

While the details can be found in various references, here we simply describe the results. We define a real $r \times r$ matrix $C(\alpha)$ by

$$C_k^j(\alpha) = c_{lk}^j \alpha^l \tag{9.57}$$

the superscript (subscript) being the row (column) index. Then the matrix $\Xi(\alpha)$ is given as an infinite power series, in fact an entire function, in $C(\alpha)$:

$$\begin{aligned} \Xi(\alpha) \equiv (\xi_k^j(\alpha)) &= \left(\frac{e^z - 1}{z}\right)_{z=C(\alpha)} \\ &= 1 + \sum_{n=1}^{\infty} (C(\alpha))^n/(n+1)! \end{aligned} \tag{9.58}$$

Because of the antisymmetry property of the structure constants, the condition reflecting the use of a canonical coordinate system is satisfied. It is in the course of developing this solution (9.58) that one sees that no more conditions need to be imposed on structure constants beyond antisymmetry and the Jacobi identities. The matrix $H(\alpha)$ inverse to $\Xi(\alpha)$ is again a power series which (in general) has a finite radius of convergence:

$$\begin{aligned} H(\alpha) &\equiv (\eta_k^j(\alpha)) \\ &= (z/(e^z - 1))_{z=C(\alpha)} \\ &= \left(-\frac{z}{2} + \frac{z}{2}\coth\frac{z}{2}\right)_{z=C(\alpha)} \end{aligned} \tag{9.59}$$

In all these expressions, the basic ingredient is the matrix $C(\alpha)$. Its intrinsic property is that acting on a vector β, it produces the Lie bracket of α with β:

$$C(\alpha)\beta = [\alpha, \beta] \tag{9.60}$$

Now that an acceptable set of functions $\eta_k^j(\alpha)$ has been constructed in canonical coordinates, it is a routine matter to solve the partial differential equations (9.12), and obtain the group composition functions $f(\alpha;\beta)$, also in canonical coordinates. The result is the so-called Baker-Campbell-Hausdorff

series. Remembering that we are using canonical coordinates, so that $\alpha^j, \beta^j, \ldots$ are both coordinates for group elements and components of tangent vectors, one finds

$$f(\alpha;\beta) = \alpha + \beta + \frac{1}{2}[\alpha,\beta] + \frac{1}{12}[\alpha-\beta,[\alpha,\beta]] + \ldots \tag{9.61}$$

The intrinsic, coordinate free form for this composition law for G reads:

$$u, v, \ldots \in \mathcal{L} \to \exp(u), \exp(v), \ldots \in G:$$
$$\exp(u)\exp(v) = \exp(u + v + \frac{1}{2}[u,v] + \frac{1}{12}[u-v,[u,v]] + \ldots) \tag{9.62}$$

In summary, we saw in sections 9.4 and 9.5 that a given Lie group G with coordinates α over a neighbourhood $\mathfrak{N}$ allows calculation of its structure constants, and thus the setting up of its Lie algebra $\underline{G}$. Conversely in this section we have indicated how, given the structure constants and working in a canonical coordinate system, we can build up $\xi(\alpha), \eta(\alpha), f(\alpha;\beta)$ for small enough α, β, and thus locally reconstruct the group.

We find in both forms (9.61),(9.62) of the group composition law that after the two leading terms, all higher order terms involve (repeated) Lie brackets. This immediately establishes the converse to a statement we made in section 9.4 (at Eq.(9.46)): if the structure constants vanish, the group is Abelian.

9.7 Comments on the $G \to \underline{G}$ Relationship

It was already mentioned in section 9.5 that the Lie algebra $\underline{G}$ associated with a Lie group G is determined by the structure of G "in the small", i.e., near the identity element. Therefore two (or more) locally isomorphic groups $G, G', \ldots$, which all "look the same" near their identities, but possibly differ globally, will share, or possess, the same Lie algebra: $\underline{G} = \underline{G'} = \ldots$. The global properties of G do not show up in $\underline{G}$ at all.

Examples familiar from nonrelativistic and relativistic quantum mechanics are: SU(2) and SO(3); SL(2,C) and SO(3,1). If then some Lie algebra $\mathcal{L}$ is given, in the reconstruction process, which of the many possible globally defined Lie groups can be uniquely singled out, as the *result* of the reconstruction? Of all the Lie groups $G, G', \ldots$ possessing the same Lie algebra $\mathcal{L}$, only *one* is simply connected, and it is called the universal covering group of all the others. For instance, SU(2) (SL(2,C)) is the universal covering group of SO(3) (SO(3,1)). Denoting this topologically distinguished group by $\bar{G}$, we can say that $\mathcal{L}$ leads unambiguously to $\bar{G}$. In other words, structure near e plus simple connectivity fixes $\bar{G}$ completely.

All of the general group theoretical notions, when specialised to Lie groups, in turn lead to corresponding notions for Lie algebras. While we look at many of them in the next section, we dispose of the simplest one now. This is the question: When are two Lie algebras $\mathcal{L}, \mathcal{L}'$ to be regarded as "the same"? How do we define *isomorphism* of Lie algebras? This requires the existence of a one-to-one, onto (hence invertible) linear map $\varphi : \mathcal{L} \to \mathcal{L}'$ (so as real vector spaces,

$\mathcal{L}$ and $\mathcal{L}'$ must have the same dimension) compatible with the two Lie bracket operations:

$$\begin{aligned} &x, y \in \mathcal{L} : \varphi(x), \varphi(y) \in \mathcal{L}'; \\ &\varphi([x, y]_{\mathcal{L}}) = [\varphi(x), \varphi(y)]_{\mathcal{L}'} \end{aligned} \tag{9.63}$$

We can then say:

$$G \text{ and } G' \textit{ locally} \text{ isomorphic} \Leftrightarrow \underline{G} \text{ and } \underline{G}' \text{ isomorphic} \tag{9.64}$$

In a similar way, homomorphisms and automorphisms for Lie algebras can easily be defined — we leave these as exercises for the reader.

9.8 Various Kinds of and Operations with Lie Algebras

We now briefly go over the notions of subalgebras, invariant, solvable, etc as applied to Lie algebras.

(i) A Lie algebra $\mathcal{L}$ is *abelian* or *commutative* if all Lie brackets vanish:

$$[x, y] = 0 \text{ for all } x, y \in \mathcal{L} \tag{9.65}$$

(ii) A subset $\mathcal{L}' \subset \mathcal{L}$ is a *subalgebra* if, firstly, it is a linear subspace of $\mathcal{L}$ viewed as a real linear vector space, and secondly, it is a Lie algebra in its own right:

$$x, y \in \mathcal{L}', a, b \in R \Rightarrow ax + by \in \mathcal{L}', [x, y] \in \mathcal{L}' \tag{9.66}$$

It is a proper subalgebra, if it is neither $\mathcal{L}$ nor the zero vector alone.

(iii) A subalgebra $\mathcal{L}' \subset \mathcal{L}$ is an *invariant* or a *normal* subalgebra (also called an *ideal*), if

$$x \in \mathcal{L}', y \in \mathcal{L} \Rightarrow [x, y] \in \mathcal{L}' \tag{9.67}$$

This can also be suggestively expressed as

$$[\mathcal{L}, \mathcal{L}'] \subset \mathcal{L}' \tag{9.68}$$

In such a case, if we project $\mathcal{L}$ as a vector space with respect to $\mathcal{L}'$, and go to the quotient $\mathcal{L}/\mathcal{L}'$ which is a vector space, it is also a Lie algebra. It is called the factor algebra. (Verify this!).

(iv) Given a Lie algebra $\mathcal{L}$, its commutator subalgebra $\mathcal{L}_1$ consists of all (real) linear combinations of Lie brackets $[x, y]$ for all choices of x, y in $\mathcal{L}$. It is also called the derived algebra of $\mathcal{L}$, and can be suggestively written as

$$\mathcal{L}_1 = [\mathcal{L}, \mathcal{L}] \tag{9.69}$$

You can easily verify that $\mathcal{L}_1$ is an invariant subalgebra of $\mathcal{L}$, and also that $\mathcal{L}/\mathcal{L}_1$ is an abelian subalgebra.

(v) Given the Lie algebra $\mathcal{L}$ of dimension r, form its commutator algebra $\mathcal{L}_1 = [\mathcal{L}, \mathcal{L}]$, then $\mathcal{L}_2 = [\mathcal{L}_1, \mathcal{L}_1]$ = commutator algebra of $\mathcal{L}_1, \mathcal{L}_3 = [\mathcal{L}_2, \mathcal{L}_2]$, and so on, $\mathcal{L}_{j+1} = [\mathcal{L}_j, \mathcal{L}_j]$ in general. In the series of Lie algebras

$$\mathcal{L}_0 \equiv \mathcal{L}, \mathcal{L}_1 = [\mathcal{L}_0, \mathcal{L}_0], \mathcal{L}_2 = [\mathcal{L}_1, \mathcal{L}_1], \ldots, \mathcal{L}_{j+1} = [\mathcal{L}_j, \mathcal{L}_j], \ldots \tag{9.70}$$

at each stage $\mathcal{L}_{j+1}$ is an invariant subalgebra of $\mathcal{L}_j$, and $\mathcal{L}_j/\mathcal{L}_{j+1}$ is abelian. If for a particular $j, \mathcal{L}_{j+1}$ is a proper subalgebra of $\mathcal{L}_j$, then the dimension of $\mathcal{L}_{j+1}$ must be at least one less than that of $\mathcal{L}_j$. On the other hand, if for some (least) j, we find $\mathcal{L}_{j+1} = \mathcal{L}_j$, then of course $\mathcal{L}_{j+1} = \mathcal{L}_{j+2} = \mathcal{L}_{j+3} = \ldots$ as well, so the series stabilises at this point. Therefore, in the series (9.70) either, (i) the dimension keeps decreasing until we reach $\mathcal{L}_j$ for some $j < r$, and thereafter the series is constant at a nonzero "value", $\mathcal{L}_j = \mathcal{L}_{j+1} = \mathcal{L}_{j+2} = \ldots \neq 0$; or (ii) the dimensions keep decreasing until, for some (least) $j \leq r, \mathcal{L}_j = 0$, and then of course $\mathcal{L}_{j+1} = \mathcal{L}_{j+2} = \ldots = 0$.

The Lie algebra $\mathcal{L}$ is *solvable* if for some (least) $j \leq r, \mathcal{L}_j = 0$.

(vi) Another sequence of subalgebras which we can form is: we go from $\mathcal{L}_0 \equiv \mathcal{L}$ to $\mathcal{L}_1$ as before, but write $\mathcal{L}^1$ for it:

$$\mathcal{L}^1 \equiv \mathcal{L}_1 = [\mathcal{L}_0, \mathcal{L}_0] \tag{9.71}$$

Next, we form

$$\begin{aligned} \mathcal{L}^2 = [\mathcal{L}_0, \mathcal{L}^1] = &\text{all (real) linear combinations of} \\ &[x, y] \text{ with } x \in \mathcal{L}_0, y \in \mathcal{L}^1 \end{aligned} \tag{9.72}$$

It is an instructive and not too complicated exercise to show that $\mathcal{L}^2$ is an invariant subalgebra of $\mathcal{L}_0$ as well as of $\mathcal{L}^1$, and $\mathcal{L}^1/\mathcal{L}^2$ is abelian. At the next step we define

$$\begin{aligned} \mathcal{L}^3 = [\mathcal{L}_0, \mathcal{L}^2] = &\text{all (real) linear combinations of} \\ &[x, y] \text{ with } x \in \mathcal{L}_0, y \in \mathcal{L}^2 \end{aligned} \tag{9.73}$$

This in turn is seen to be an invariant subalgebra of $\mathcal{L}_0, \mathcal{L}^1$ and $\mathcal{L}^2$, with $\mathcal{L}^2/\mathcal{L}^3$ being abelian. In this way we have the series:

$$\begin{aligned} \mathcal{L}_0 \equiv \mathcal{L}, \mathcal{L}^1 \equiv \mathcal{L}_1 = [\mathcal{L}_0, \mathcal{L}_0], \mathcal{L}^2 = [\mathcal{L}_0, \mathcal{L}^1], \ldots, \\ \mathcal{L}^{j+1} = [\mathcal{L}_0, \mathcal{L}^j], \ldots \end{aligned} \tag{9.74}$$

$\mathcal{L}^{j+1}$ is an invariant subalgebra of $\mathcal{L}_0, \mathcal{L}^1, \mathcal{L}^2, \ldots, \mathcal{L}^j$ and $\mathcal{L}^j/\mathcal{L}^{j+1}$ is abelian.

The Lie algebra $\mathcal{L}$ is *nilpotent* if $\mathcal{L}^j = 0$ for some (least) $j \leq r$.

(vii) Connection between solvability and nilpotency: We have the two series (9.70),(9.74) of Lie algebras above. While $\mathcal{L}_1 = \mathcal{L}^1$ by definition, one can easily show (by induction for example) that

$$\mathcal{L}_j \subseteq \mathcal{L}^j, j = 2, 3, \ldots \tag{9.75}$$

This means that if $\mathcal{L}^j = 0$ for some j, then by that stage (or earlier) $\mathcal{L}_j = 0$ too. So,

$$\mathcal{L} \text{ nilpotent} \Rightarrow \mathcal{L} \text{ solvable, but not conversely} \tag{9.76}$$

While the above concepts are important for the general theory of Lie algebras, for most physical applications the simple and semisimple types are more important. We next define such Lie algebras.

(viii) A Lie algebra $\mathcal{L}$ is *simple (semisimple)* if it has no proper invariant (abelian invariant) subalgebra. We can immediately see the following relationships:

$$\begin{aligned} \mathcal{L} \text{ simple} &\Rightarrow \mathcal{L} \text{ semisimple but not conversely;} \\ \mathcal{L} \text{ solvable} &\Rightarrow \mathcal{L}_1 = \text{ proper invariant subalgebra of } \mathcal{L} \\ &\Rightarrow \mathcal{L} \text{ not simple;} \\ \mathcal{L} \text{ simple} &\Rightarrow \mathcal{L}_1 = \mathcal{L} \Rightarrow \mathcal{L} \text{ not solvable} \end{aligned} \tag{9.77}$$

So, simplicity and solvability are mutually exclusive! To conclude this section, let us define direct and semidirect sums of Lie algebras.

(ix) Direct sum of Lie algebras: Let $\mathcal{L}'$ be a proper invariant subalgebra of a Lie algebra $\mathcal{L}$. If we can find another proper invariant subalgebra $\mathcal{L}''$ in $\mathcal{L}$ such that as vector spaces $\mathcal{L} = \mathcal{L}' \oplus \mathcal{L}''$ in the sense of a direct sum, then we say $\mathcal{L}$ is the *direct sum Lie algebra* of $\mathcal{L}'$ and $\mathcal{L}''$. Since in this situation the only element common to $\mathcal{L}'$ and $\mathcal{L}''$ is the zero element, we must have:

$$[\mathcal{L}', \mathcal{L}'] \subseteq \mathcal{L}', [\mathcal{L}', \mathcal{L}''] = 0, [\mathcal{L}'', \mathcal{L}''] \subseteq \mathcal{L}'' \tag{9.78}$$

(x) The semidirect sum of Lie algebras parallels the semidirect product of groups. We say $\mathcal{L}$ is the semidirect sum of $\mathcal{L}'$ and $\mathcal{L}''$ if as vector spaces $\mathcal{L} = \mathcal{L}' \oplus \mathcal{L}''$, and as Lie algebras

$$\begin{aligned} [\mathcal{L}', \mathcal{L}'] &\subseteq \mathcal{L}', \\ [\mathcal{L}', \mathcal{L}''] &\subseteq \mathcal{L}'', \\ [\mathcal{L}'', \mathcal{L}''] &\subseteq \mathcal{L}'', \end{aligned} \tag{9.79}$$

This means that while both $\mathcal{L}'$ and $\mathcal{L}''$ are subalgebras of $\mathcal{L}$, the latter is an invariant one; and one can then see that the factor algebra $\mathcal{L}/\mathcal{L}''$ is isomorphic to $\mathcal{L}'$. From the second line in Eq.(9.79) we also see that each element of $\mathcal{L}'$ acts as an (in general) outer automorphism on $\mathcal{L}''$.

Exercises for Chapter 9

1. For the group SO(3) of real proper orthogonal rotations in 3 dimensions, show that we have various possible parametrisations of group elements:

(i) Euler angles:

$$R \in \mathrm{SO}(3) : R = R(\psi, \theta, \phi) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix},$$
$$0 \leq \theta, \phi \leq 2\pi, 0 \leq \psi \leq \pi.$$

(ii) Axis angle parameters:

$$R(\hat{n}, \alpha) = (R_{jk}(\hat{n}, \alpha)), \quad j, k = 1, 2, 3,$$
$$R_{jk}(\hat{n}, \alpha) = \delta_{jk} \cos\alpha + n_j n_k (1 - \cos\alpha) - \epsilon_{jkl} n_l \sin\alpha,$$
$$\hat{n} \in S^2, 0 \leq \alpha \leq \pi.$$

(iii) Homogeneous Euler parameters:

$$R_{jk}(a_0, \underline{a}) = (a_0^2 - \underline{a}^2)\delta_{jk} + 2a_j a_k - 2a_0 \epsilon_{jkl} a_l,$$
$$a_0 = \cos\frac{\alpha}{2}, a_j = n_j \sin\frac{\alpha}{2}, a_0^2 + \underline{a}^2 = 1$$

In each case, find the parameter values for the identity element. Are they always unique?

2. For the group SO(3) described using axis angle parameters $R(\hat{n}, \alpha)$, show that by keeping $\hat{n}$ fixed we obtain a one-parameter subgroup. What is the range of α?

3. Develop analogues of (i), (ii), (iii) of problem (1) above for the group SU(2). Find all possible one-parameter subgroups in this case.

Chapter 10

Linear Representations of Lie Algebras

We have reviewed in Chapter 8 some aspects of the theory of group representations. When applied to Lie groups in particular, they lead to linear representations of Lie algebras. We devote this relatively short Chapter to this topic.

Let a Lie group G and its Lie algebra (sometimes written as $\underline{G}$ and some times in a general context as $\mathcal{L}$) be both given. To get a linear representation of $\mathcal{L}$, we must start with a (real or complex) linear vector space $\mathcal{V}$, of dimension n say, and set up the following association:

$$\begin{aligned} &\text{Elements in } \mathcal{L} \rightarrow \text{linear transformations on } \mathcal{V}, \\ &\text{Lie bracket in } \mathcal{L} \rightarrow \text{commutator of transformations on } \mathcal{V} \end{aligned} \tag{10.1}$$

Why is the abstract Lie bracket in $\mathcal{L}$ "realised" as the commutator of linear operators in a linear representation? Go back for a moment to the Lie group G: actually the ensuing operations refer only to a suitable neighbourhood $\mathfrak{N} \subset G$ containing the identity. If we had a representation of G on $\mathcal{V}$, then to the element $a \in \mathfrak{N}$ with coordinates α we would have the representation matrix $D(\alpha)$ say, obeying:

$$\begin{aligned} D(\alpha)D(\beta) &= D(f(\alpha;\beta)) \\ D(0) &= 1 \end{aligned} \tag{10.2}$$

What would the elements on an OPS look like in the representation? Assume that the coordinate system is a canonical one, and for small α, suppose

$$D(\alpha) \simeq 1 + \alpha^j X_j + 0(\alpha^2) \tag{10.3}$$

Thus, given a basis for $\mathcal{L}$, $\{e_j\}$ say, the way the α's are enumerated is fixed: and in the representation on $\mathcal{V}$ we find that to each e_j there corresponds a linear operator X_j on $\mathcal{V}$:

$$e_j \in \mathcal{L} \rightarrow \text{ linear operator } X_j \text{ on } \mathcal{V} \tag{10.4}$$

These are called the *generators* of the representation. Next, for two elements $a, b \in \mathfrak{N}$, both close to e, the commutator group element $q(a,b)$ has coordinates given by Eq.(9.46). Therefore we have the condition,

$$\begin{aligned} D(q(a,b)) &\equiv D(\alpha)D(\beta)D(-\alpha)D(-\beta) \\ &\simeq 1 + c^j_{kl}\alpha^k\beta^l X_j + \dots \end{aligned} \tag{10.5}$$

If we put in expressions like (10.3) on the left hand side and retain terms at most linear in α and β, we immediately find:

$$\alpha^k X_k \beta^l X_l - \beta^l X_l \alpha^k X_k = c^j_{kl}\alpha^k\beta^l X_j,$$

$$\text{i.e.,} \quad [X_k, X_l] \equiv X_k X_l - X_l X_k = c^j_{kl} X_j \tag{10.6}$$

And then the elements on an OPS are of course realised as "ordinary exponentials":

$$\begin{aligned} D(\exp(u)) &= 1 + \frac{u\cdot X}{1!} + \frac{(u\cdot X)^2}{2!} + \dots \\ &= e^{u\cdot X} \\ u\cdot X &= u^j X_j \end{aligned} \tag{10.7}$$

One sees how the up-to-now formal exponential used in relating vectors in $\mathcal{L}$ to elements in G via OPS's becomes the ordinary exponential in a linear representation. In other words, if we have a linear representation of $\mathcal{L}$ on $\mathcal{V}$, with the Lie bracket of $\mathcal{L}$ interpreted as the ordinary commutator of generators acting on $\mathcal{V}$, then (apart from global aspects) upon exponentiation we get a representation of G: in fact since every $a \in \mathfrak{N}$ does lie on some OPS, the result (10.7) does fix the way all elements in $\mathfrak{N} \subset G$ are to be represented.

(A more precise statement about global aspects is this. A hermitian representation of the Lie algebra $\mathcal{L}$ of a Lie group G yields, upon exponentiation a true representation of $\tilde{G}$, the universal covering group of G, and in general not of G itself).

This situation can be "turned around" in the following sense. Many groups of practical interest are defined via a specific linear representation on some definite space $\mathcal{V}$ (a defining matrix representation), which is declared to be globally faithful. Then we can in principle "look" at the family of matrices involved, find independent real parameters (at least locally) for them, evaluate the generator matrices in that representation, and from their commutation properties "read off" the structure constants needed to search for other representations! In fact this is the most practically convenient way to handle the orthogonal, unitary and symplectic groups, as we will see later.

At this point, a matter of convention and notation needs to be explained. In quantum mechanical applications unitary representations of groups play an important role. It is clear from Eq.(10.7) that for $D(a)$ to be *unitary*, (assuming $\mathcal{V}$ is equipped with the appropriate inner product), the generators X_j must be *antihermitian*. It is however usual to remove explicitly a factor of i and deal

with *hermitian generators*. Hereafter we shall uniformly adhere to this quantum mechanical convention. In the context of unitary (hermitian) representations of $G(\mathcal{L})$, we shall instead of the previous Eqs.(10.3),(10.6),(10.7) adopt the following practice:

$$\text{Unitary representation of } G \leftrightarrow \text{Hermitian representation of } \mathcal{L} \qquad \text{(a)}$$

$$D(\alpha) \simeq 1 - i\alpha^j X_j + 0(\alpha^2); \qquad \text{(b)}$$

$$e_j \rightarrow X_j = X_j^\dagger; \qquad \text{(c)}$$

$$[e_j, e_k] = c_{jk}^l e_l \rightarrow [X_j, X_k] = ic_{jk}^l X_l \qquad \text{(d)}$$

$$a = \exp(u^j e_j) \rightarrow D(a) = \exp(-iu^j X_j) \qquad \text{(e)}$$

(10.8)

Alternatively we could say that the abstract Lie bracket in $\mathcal{L}$ goes into $-i$ times the commutator.

A real orthogonal representation of G (on a real $\mathcal{V}$ with appropriate inner product) leads to generator matrices X_j which are hermitian and *antisymmetric*, hence purely imaginary. To emphasize that a common quantum mechanical convention is being used, we express the situation thus:

$$\text{Unitary group representation} \leftrightarrow \text{hermitian generators},$$
$$X_j = X_j^\dagger;$$
$$\text{Real orthogonal group representations} \leftrightarrow \text{hermitian antisymmetric generators},$$
$$X_j^\dagger = X_j, \quad X_j^T = -X_j; \qquad (10.9)$$

Real representation $\leftrightarrow$ pure imaginary generators, $X_j^* = -X_j$.

In fact, even for non unitary non-real orthogonal representations, the same conventions (10.8) apply. In this most general case, X_j are neither hermitian nor antisymmetric imaginary.

For Lie algebra representations, all the notions of invariant subspaces, irreducibility, reducibility, decomposability, direct sum etc can be directly carried over from the discussions of Chapter 8. Other than these, the passages to the contragredient, adjoint, and complex conjugate group representations are reflected at the generator level thus:

$$D \rightarrow (D^T)^{-1} : X_j \rightarrow -X_j^T;$$
$$D \rightarrow (D^\dagger)^{-1} : X_j \rightarrow X_j^\dagger;$$
$$D \rightarrow D^* : X_j \rightarrow -X_j^* \qquad (10.10)$$

In these rules, the "quantum mechanical i" has been properly respected.

We finally take up in this Chapter a particular real representation of a Lie algebra $\mathcal{L}$ and "its Lie group G" (locally), which is intrinsic to the structure of

$\mathcal{L}$ and is called the *adjoint representation.* The beauty of this representation is that $\mathcal{L}$ itself serves as the representation space $\mathcal{V}$, on which both $\mathcal{L}$ and G are made to act! For each $x \in \mathcal{L}$, define a linear transformation ad $\cdot x$ which acts on $\mathcal{L}$ in this way:

$$\begin{aligned} &(\text{ad. } x) : y \to y' = [x, y], \\ &\text{i.e.,} \qquad (\text{ad. } x)y = [x, y] \end{aligned} \tag{10.11}$$

Here the abstract Lie bracket in $\mathcal{L}$ has been used, and the linearity properties of this bracket ensure that ad. x is a linear transformation. Do we have here a representation of $\mathcal{L}$? Yes, by Jacobi!

$$\begin{aligned} (\text{ad. } x)(\text{ad. } y)z - (\text{ad. } y)(\text{ad. } x)z &= [x, [y, z]] - [y, [x, z]] \\ &= [[x, y], z] \\ &= (\text{ad. } [x, y])z, \\ [\text{ad. } x, \text{ad. } y] &= \text{ad. } [x, y] \end{aligned} \tag{10.12}$$

(The brackets on the two sides of this last line are formally different!). In a basis $\{e_j\}$ for $\mathcal{L}$, the generator matrices are the structure constants (times a factor of i):

$$\begin{aligned} &(\text{ad. } e_j)e_k = [e_j, e_k] = c^l_{jk} e_l \Rightarrow \\ &(X^{(\text{adj})}_j)^l_k = i \times \text{ coefficient of } e_l \text{ in } (\text{ad. } e_j)e_k \\ &\qquad\qquad = i c^l_{jk} \end{aligned} \tag{10.13}$$

Then the Jacobi identity (9.53) for the structure constants amounts to the generator commutation relations:

$$[X^{(\text{adj})}_j, X^{(\text{adj})}_k] = i c^l_{jk} X^{(\text{adj})}_l \tag{10.14}$$

In the expressions in Eq.(10.13), while one of the subscripts on the structure constant is used up to enumerate the X's, the remaining superscript (subscript) acts as a matrix row (column) index.

Exponentiation of the adjoint representation of $\mathcal{L}$ leads of course to a real local representation of G on $\mathcal{L}$ itself. This, in its global aspect, can be called the adjoint representation of the simply connected universal covering group $\bar{G}$, which $\mathcal{L}$ determines unambiguously. Alternatively, one also often defines as the *adjoint group* of $\mathcal{L}$ that Lie group which possesses $\mathcal{L}$ as its Lie algebra, and for which by definition the set of all exponentials of the generators $iu^j X^{(\text{adj})}_j$, and products thereof, give a faithful representation.

The adjoint representation will play an important role in all our considerations.

Exercises for Chapter 10

1. Show that among functions $f(\underline{q},\underline{p}), g(\underline{q},\underline{p}), \ldots$ on a classical mechanical real $2n$-dimensional phase space, the Poisson bracket

$$\{f,g\} = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial q_j}\frac{\partial g}{\partial p_j} - \frac{\partial f}{\partial p_j}\frac{\partial g}{\partial q_j} \right)$$

   is a realisation of the Lie bracket concept. Similarly show that among three-dimensional real vectors $\underline{a},\underline{b},\ldots$, the vector product operation

$$\underline{a},\underline{b} \longrightarrow \underline{a} \wedge \underline{b}$$

   is also a realisation of this concept.

2. From the quantum theory of angular momentum we know that the Lie algebra commutation relations for generators of SO(3) are

$$[J_j, J_k] = i\epsilon_{jkl} J_l,$$

   ϵ_{jkl} = completely antisymmetric, $\epsilon_{123} = 1$. Using Eq.(10.13), verify that the adjoint representation of SO(3) is its real 3 dimensional defining representation.

Chapter 11

Complexification and Classification of Lie Algebras

So far we have dealt with real (finite dimensional) Lie algebras, as they arise while analysing Lie groups whose elements can be parametrised with an essential and finite number of real independent coordinates. The task of trying to systematically classify all such Lie algebras is a very difficult and complex one, and it would definitely be out of place to attempt to describe here in all detail the results known in this area. On the other hand, it seems worthwhile at least introducing some of the key concepts that are used in this branch of mathematics, and conveying the flavour of the subject. We attempt to do no more than this in the present Chapter. To go further, the reader is advised to study one of the works in the list of references given at the end.

We describe the process of complexification of a real Lie algebra, and give the definition of a complex Lie algebra. We also indicate how the properties of solvability, semisimplicity and simplicity are used in the classification programme. Our final aim is to arrive at and deal with the series of Compact Simple Lie Algebras (CSLA), and their associated Lie groups.

11.1 Complexification of a Real Lie Algebra

Let $\mathcal{L}$ be a real r-dimensional Lie algebra. By a straightforward procedure we can complexify it and arrive at a complex Lie algebra $\tilde{\mathcal{L}}$. To begin, we complexify $\mathcal{L}$ as a vector space. Thus the complex r-dimensional vector space $\tilde{\mathcal{L}}$ consists of formal expressions of the form,

$$\begin{aligned} z = x + iy \in \tilde{\mathcal{L}}, \\ x, y \in \mathcal{L} \end{aligned} \tag{11.1}$$

Two such expressions are the same only if the separate parts are equal:

$$x + iy = x' + iy' \Leftrightarrow x = x', y = y' \tag{11.2}$$

For a complex number $a + ib$, with a and b real, we define

$$(a + ib)(x + iy) = ax - by + i(ay + bx) \tag{11.3}$$

which is again of the form (11.1). You can check that

$$(a + ib)(x + iy) = 0 \Leftrightarrow \text{ either } a + ib = 0 \text{ or } x + iy = 0 \text{ or both} \tag{11.4}$$

Finally, then, Lie brackets in $\tilde{\mathcal{L}}$ are defined by

$$[x + iy, u + iv] = [x, u] - [y, v] + i([x, v] + [y, u]) \tag{11.5}$$

The validity of the Jacobi identity is obvious.

We say that $\tilde{\mathcal{L}}$ is the unique complex form of the real Lie algebra $\mathcal{L}$. The complex dimension of $\tilde{\mathcal{L}}$ is the same as the real dimension of $\mathcal{L}$; and a basis $\{e_j\}$ for $\mathcal{L}$ remains a basis for $\tilde{\mathcal{L}}$ as well. In this basis, the structure constants of $\tilde{\mathcal{L}}$ are the same as those of $\mathcal{L}$, namely c^l_{jk}, and so they are real.

Having seen how $\tilde{\mathcal{L}}$ arises from $\mathcal{L}$, we can now directly define what we mean by a complex Lie algebra. A complex Lie algebra $\tilde{\mathcal{L}}$ of dimension r is a complex r-dimensional vector space, with elements $z, z', \ldots \in \tilde{\mathcal{L}}$, among which a Lie bracket is defined:

$$\begin{aligned} z, z' \in \tilde{\mathcal{L}}, \lambda, \mu \in \mathbb{C}; \\ \lambda z + \mu z' \in \tilde{\mathcal{L}}; \\ [z, z'] \in \tilde{\mathcal{L}}; \\ [z, z'] = \text{ linear, antisymmetric;} \\ [[z, z'], z''] + [[z', z''], z] + [[z'', z], z'] = 0 \end{aligned} \tag{11.6}$$

This idea of a complex Lie algebra is directly defined in this way, and not derived from a finite dimensional Lie group with real parameters. If $\{\tilde{e_j}\}$ is a basis for $\tilde{\mathcal{L}}$, in general we have complex structure constants $\tilde{c}^l_{jk}$ obeying antisymmetry and the Jacobi condition.

For a general complex Lie algebra, there may be no basis in which all the structure constants become real! Of course an r-dimensional complex Lie algebra $\tilde{\mathcal{L}}$ can be viewed as a real $2r$-dimensional one with an additional operation representing "multiplication by i".

Now let us mention some statements which are intuitively evident, partly repeating what was said earlier:

(i) A given real r-dimensional Lie algebra $\mathcal{L}$ leads to a unique complex r-dimensional Lie algebra $\tilde{\mathcal{L}}$, its *complex form*. A basis for $\mathcal{L}$ can be used as a basis for $\tilde{\mathcal{L}}$, and in that case the structure constants are real and unchanged.

(ii) Many distinct real Lie algebras $\mathcal{L}, \mathcal{L}', \ldots$ may lead to the same complex one $\tilde{\mathcal{L}}$; in that case each of the former is called a *real form* of the latter, and $\tilde{\mathcal{L}}$ arises by complexification of any of $\mathcal{L}, \mathcal{L}', \ldots$.

(iii) A given complex r-dimensional Lie algebra $\tilde{\mathcal{L}}$ may have none, or many, real forms. If it has *none*, then there is *no* basis for $\tilde{\mathcal{L}}$ with respect to which all the structure constants have real values!

While the use of complex Lie algebras is useful and indeed unavoidable for purposes of the general theory, the process of "returning to the real forms" is made quite difficult because of the above facts. As we will see later, though, some fundamental theorems of Weyl help simplify the situation for the real compact simple Lie algebras.

The definitions of subalgebras, invariant subalgebras, factor algebras etc can all be given quite easily for complex Lie algebras. The relation between such properties for a real $\mathcal{L}$ and for its complex form $\tilde{\mathcal{L}}$ is this:

$$\begin{aligned} \mathcal{L} \text{ abelian} &\Leftrightarrow \tilde{\mathcal{L}} \text{ abelian;} \\ \mathcal{L} \text{ solvable} &\Leftrightarrow \tilde{\mathcal{L}} \text{ solvable;} \\ \mathcal{L} \text{ simple} &\not\Leftrightarrow \tilde{\mathcal{L}} \text{ simple} \end{aligned} \tag{11.7}$$

We have a surprise in the last line! The "fly in the ointment" is the case of the groups $SL(n,C)$; while the real Lie algebra $\underline{SL(n,C)}$ is simple, after complexification it is no longer simple.

11.2 Solvability, Levi's Theorem, and Cartan's Analysis of Complex (Semi) Simple Lie Algebras

It turns out that solvability is a crucial generalisation of abelianness, and it has a hereditary and contagious character. For both real and complex cases one finds solvability of a Lie algebra implies the same for all its subalgebras and for the factor algebras with respect to its invariant subalgebras. Even more is true: if a (real or complex) Lie algebra $\mathcal{L}$ has an invariant subalgebra $\mathcal{L}'$, and both $\mathcal{L}'$ and the factor $\mathcal{L}/\mathcal{L}'$ are solvable, then so is $\mathcal{L}$ itself!

Levi Splitting Theorem

Any (real or complex) Lie algebra $\mathcal{L}$ can be split into the semidirect sum of a maximal solvable invariant subalgebra S and a semisimple subalgebra T:

$$\begin{aligned} L &= S+)T, \\ [S,S], [T,S] &\subseteq S, \\ [T,T] &\subseteq T \end{aligned} \tag{11.8}$$

Based on this result, can we classify all real or complex Lie algebras? No! while the semisimple ones can be classified, this has so far not been possible for the solvable ones. However having seen this basic theorem, and appreciated where the situation regarding classification now stands, we shall hereafter consider mainly the semisimple case.

The classification of complex semisimple Lie algebras rests heavily on some fundamental theorems of Cartan, and work and analysis by Killing, Cartan and others. We give a qualitative account of these matters, staying at the complex level up to a certain point, then switching to the real compact simple cases. We begin with

Theorem (Cartan, 1894)

A (real or complex) semi-simple Lie algebra $\mathcal{L}$ is the direct sum of mutually commuting simple nonabelian subalgebras:

$$\begin{aligned} &\mathcal{L} \text{ semisimple} \Leftrightarrow \mathcal{L} = \mathcal{L}' \oplus \mathcal{L}'' \oplus \ldots, \quad \mathcal{L}', \mathcal{L}'', \ldots \text{ simple, nonabelian,} \\ &[\mathcal{L}', \mathcal{L}''] = [\mathcal{L}', \mathcal{L}'''] = [\mathcal{L}'', \mathcal{L}'''] = \ldots = 0 \end{aligned} \tag{11.9}$$

Next, as a result of the work of Killing, Cartan and others, the *complex simple Lie algebras have all been found and classified.* There are four infinite families and five exceptional cases. We will deal later with the compact real forms of these. For the present, let us see how the structure of complex simple Lie algebras has been analysed. The method rests largely on skillful exploitation of the Jacobi identities for structure constants. Incidentally we can always keep in mind that analysis of the adjoint representation amounts to analysis of the basic commutation relations and so of structure constants.

To be reasonably well organized, let us number the statements to follow, and arrange them in as good a sequence as we can!

1. Let the (complex r-dimensional) Lie algebra $\tilde{\mathcal{L}}$ in some basis have structure constants c_{jk}^m which could be complex (we save the index l for a good reason); and define a "metric tensor"

$$g_{jk} = g_{kj} = c_{jn}^m c_{km}^n, \tag{11.10}$$

the sums on m and n being understood. For simplicity we avoid tildes on c and g, though they may be complex. A result of Cartan says:

$$\tilde{\mathcal{L}} \text{ is semisimple } \Leftrightarrow |g_{jk}| = \det(g_{jk}) \neq 0 \tag{11.11}$$

In one direction, the proof is very easy. If $\tilde{\mathcal{L}}$ has an abelian invariant subalgebra, we can choose a basis in $\tilde{\mathcal{L}}$ such that certain rows and columns of (g_{jk}) vanish identically, so

$$\begin{aligned} &\tilde{\mathcal{L}} \text{ not semisimple } \Rightarrow |g| = 0, \\ &\qquad |g| \neq 0 \Rightarrow \tilde{\mathcal{L}} \text{ semisimple} \end{aligned} \tag{11.12}$$

The harder thing to show is that

$$\begin{aligned} \tilde{\mathcal{L}} \text{ semisimple } &\Rightarrow |g| \neq 0, \\ |g| = 0 &\Rightarrow \tilde{\mathcal{L}} \text{ not semisimple} \end{aligned} \tag{11.13}$$

We leave this as an exercise to the interested reader. In particular, Eq.(11.11) gives:

$$\tilde{\mathcal{L}} \text{ simple } \Rightarrow |g_{jk}| \neq 0. \tag{11.14}$$

Hereafter, we assume $\tilde{\mathcal{L}}$ is simple, so we can and will use the fact that (g) is nonsingular.

2. From the structure constants c^m_{jk}, by lowering the superscript using the metric, we can construct a three-subscript symbol which is totally antisymmetric:

$$c_{jkm} = g_{mn}c^n_{jk} \tag{11.15}$$

Since (g) is nonsingular, we can always get back the true structure constants from here. The antisymmetry in the first two indices is evident. The Jacobi identity leads to antisymmetry in the second and third:

$$\begin{aligned} c_{jkm} &= c^n_{jk}c^p_{nq}c^q_{mp} && \text{(a)} \\ &= -(c^n_{kq}c^p_{nj} + c^n_{qj}c^p_{nk})c^q_{mp} \\ &= -c^n_{kq}c^q_{mp}c^p_{nj} - c^q_{mp}c^p_{kn}c^n_{jq} && \text{(b)} \\ &= c^n_{kq}c^q_{mp}c^p_{jn} - c^n_{mq}c^q_{kp}c^p_{jn} \\ &= (c^n_{kq}c^q_{mp} - c^n_{mq}c^q_{kp})c^p_{jn}. \end{aligned} \tag{11.16}$$

Here, at the second step (a), we have used the Jacobi identity for the first two c's at the previous step; and at the fourth step (b) in the second term the dummy indices were changed according to $p \to q \to n \to p$. Thus we have established

$$c_{jkm} = -c_{jmk} \tag{11.17}$$

and the total antisymmetry of c_{jkm} follows. We will use this later on at point no. 7.

3. Now choose an element $A \in \tilde{\mathcal{L}}$ and set up the "eigenvalue problem"

$$\begin{aligned} (\text{ad. } A)X \equiv [A, X] &= \rho X, \\ X \in \tilde{\mathcal{L}}, \rho &\in \mathbb{C}. \end{aligned} \tag{11.18}$$

One solution is of course $\rho = 0, X = A$. In any case, the number of distinct eigenvalues or roots can be no more than r-dimension of $\tilde{\mathcal{L}}$. Now vary A and choose it so as to maximise the number of distinct roots ρ. Then another Theorem of Cartan says that for such an A,

(a) There is a set $\mathfrak{R}_0$ of distinct *non-zero* roots $\alpha, \beta, \ldots$

(b) Each of these is *nondegenerate*, i.e., for each $\alpha \in \mathfrak{R}_0$, there is a *unique* $E_\alpha \in \tilde{\mathcal{L}}$ (unique upto scalar multiplication) obeying

$$\alpha \in \mathfrak{R}_0 : [A, E_\alpha] = \alpha E_\alpha. \tag{11.19}$$

(c) Only the root $\rho = 0$ is degenerate, and the degree of degeneracy l of this root is characteristic of $\tilde{\mathcal{L}}$ and is called the *rank* of $\tilde{\mathcal{L}}$.

(d) The corresponding l "eigenvectors" are elements $H_a \in \tilde{\mathcal{L}}, a = 1, 2, \ldots, l$; they are linearly independent and obey

$$[H_a, H_b] = 0, \quad a, b = 1, 2, \ldots, l \tag{11.20}$$

(e) The H_a for $a = 1, 2, \ldots, l$ and E_α for $\alpha \in \mathfrak{R}_0$ are independent and give a basis for $\tilde{\mathcal{L}}$.

It must be clear that a "maximal" A for which all this happens is by no means unique! At this point, we see that $\mathfrak{R}_0$ consists of $r - l$ distinct nonzero (complex) numbers. We will soon refine and get a better description of $\mathfrak{R}_0$.

4. Some immediate consequences, on exploiting the Jacobi identities, follow. Since A is a solution to (11.18) for $\rho = 0$, it is a linear combination of the H_a:

$$\begin{aligned} &A = \lambda^a H_a, \\ &\lambda^a \in \mathbb{C}, \text{ not all identically zero} \end{aligned} \tag{11.21}$$

Then, using the three-term Jacobi identity for A, H_a and E_α, we get

$$[A, [H_a, E_\alpha]] = \alpha [H_a, E_\alpha],$$

$$\text{i.e.} \qquad [H_a, E_\alpha] = \alpha_a E_\alpha, \qquad \alpha_a \in \mathbb{C} \tag{11.22}$$

Thus, for each (complex) number $\alpha \in \mathfrak{R}_0$, there is a (complex) l-component "root vector" $\{\alpha_a\}$, and since from Eqs.(11.19),(11.21),(11.22), we get

$$\alpha = \lambda^a \alpha_a, \tag{11.23}$$

each vector $\{\alpha_a\}$ cannot vanish identically. In fact we can say:

$$\begin{aligned} &\alpha \in \mathfrak{R}_0 \Rightarrow \{\alpha_a\} \neq 0; \\ &\alpha, \beta \in \mathfrak{R}_0, \alpha \neq \beta \Rightarrow \{\alpha_a\} \neq \{\beta_a\}. \end{aligned} \tag{11.24}$$

The set $\mathfrak{R}_0$ consists of $r - l$ distinct non-zero (complex) *numbers*. Let us now define $\mathfrak{R}$ to be the set of the corresponding $r - l$ distinct non-vanishing (complex) *root vectors* $\underline{\alpha}, \underline{\beta}, \ldots$ with components $\alpha_a, \beta_a, \ldots$. Elements of $\mathfrak{R}_0$ and of $\mathfrak{R}$ are connected by Eq.(11.23). We are of course free to replace the H_a by nonsingular linear combinations of themselves, leaving the E_α and the original A unchanged. If we do this we will regard the vectors $\underline{\alpha}, \underline{\beta}, \ldots \in \mathfrak{R}$ as not having changed, but as being resolved in a new basis.

5. The information so far collected on the structure constants c_{jk}^m is this: each index $j, k, m, \ldots = 1, 2, \ldots, r$ goes partly over $a, b, \ldots = 1, 2, \ldots, l$, and partly over the $r - l$ distinct non-zero numbers $\alpha \in \mathfrak{R}_0$ (or if we wish, $\underline{\alpha} \in \mathfrak{R}$). This is the result of using H_a and E_α as a basis for $\tilde{\mathcal{L}}$. Then,

(a) Property (3d) above $\Rightarrow c_{ab}^m = 0$ (11.25)

(b) Property (4) above $\Rightarrow c^m_{a\alpha} = \alpha_a \delta_{m,\alpha}$ (11.26)

(c) If $\alpha, \beta, \alpha+\beta \in \mathfrak{R}_0$, use of the Jacobi Identity on A, E_α, E_β gives

$$\begin{aligned} [E_\alpha, E_\beta] &= N_{\alpha,\beta} E_{\alpha+\beta}, \\ N_{\alpha,\beta} &= -N_{\beta,\alpha}, \\ c^m_{\alpha\beta} &= N_{\alpha,\beta}\delta_{m,\alpha+\beta} \end{aligned} \tag{11.27}$$

(d) If $\alpha, \beta \in \mathfrak{R}_0, \alpha+\beta \neq 0, \alpha+\beta \notin \mathfrak{R}_0$, then necessarily

$$\begin{aligned} [E_\alpha, E_\beta] &= 0, \\ c^m_{\alpha\beta} &= 0 \end{aligned} \tag{11.28}$$

It may be mentioned that (c) and (d) above are necessary consequences of the Jacobi identities, in advance of knowing the contents of the set $\mathfrak{R}_0$.

6. Now we will show that $\alpha \in \mathfrak{R}_0$ implies $-\alpha \in \mathfrak{R}_0$ as well. In turn, and this is nontrivial, it will mean that $\underline{\alpha} \in \mathfrak{R}$ implies $-\underline{\alpha} \in \mathfrak{R}$ too. Consider the subset of $(r-l)$ rows of the matrix (g_{jk}) in which $j = \alpha \in \mathfrak{R}_0$. Indicating all summations explicitly, we have, using results gathered above:

$$\begin{aligned} g_{\alpha k} &= \sum_{j,m} c^j_{\alpha m} c^m_{kj} \\ &= \sum_j \left(\sum_{a=1}^{l} c^j_{\alpha a} c^a_{kj} + \sum_{\beta \in \mathfrak{R}_0} c^j_{\alpha\beta} c^\beta_{kj} \right) \\ &= \sum_j \left(-\sum_{a=1}^{l} \alpha_a \delta_{j,\alpha} c^a_{kj} + \sum_{\substack{\beta \in \mathfrak{R}_0 \\ (\alpha+\beta \in \mathfrak{R}_0)}} N_{\alpha,\beta} \delta_{j,\alpha+\beta} c^\beta_{kj} + (c^j_{\alpha,-\alpha} c^{-\alpha}_{kj}) \right) \\ &= -\sum_{a=1}^{l} \alpha_a c^a_{k\alpha} \sum_{\substack{\beta \in \mathfrak{R}_0 \\ (\alpha+\beta \in \mathfrak{R}_0)}} N_{\alpha\beta} c^\beta_{k,\alpha+\beta} + \sum_j (c^j_{\alpha,-\alpha} c^{-\alpha}_{kj}) \end{aligned} \tag{11.29}$$

In the preceding two lines we have put inside parentheses a term which is present only if $-\alpha \in \mathfrak{R}_0$, not otherwise. If indeed $-\alpha \in \mathfrak{R}_0$, then the Jacobi identity for A, E_α and $E_{-\alpha}$ gives

$$\begin{aligned} [E_\alpha, E_{-\alpha}] &= \alpha^a H_a, \ \alpha^a \in \mathbb{C} \\ c^j_{\alpha,-\alpha} &= \alpha^a \qquad \text{if } j = a, \\ &= 0 \qquad \text{if } j = \beta \in \mathfrak{R}_0 \end{aligned} \tag{11.30}$$

Now going back to Eq.(11.29), we can take either $k = b = 1, 2, \ldots, l$ or $k = \gamma \in \mathfrak{R}_0$. In each case we can list contributions from each of the three terms, assuming in the latter case that $\gamma \neq -\alpha$:

$$\begin{aligned} k = b, \ g_{\alpha b} &= 0 + 0 + (0), \\ k = \gamma \neq -\alpha, \ g_{\alpha\gamma} &= 0 + 0 + (0). \end{aligned} \tag{11.31}$$

Here, we have used Eqs.(11.25),(11.26),(11.30) when necessary. We can conclude that

$$\begin{aligned} \alpha \in \mathfrak{R}_0, -\alpha \notin \mathfrak{R}_0 \Rightarrow g_{\alpha k} = 0 \quad \text{for all } k \\ \Rightarrow |g_{jk}| = 0, \end{aligned} \tag{11.32}$$

which contradicts Cartan's theorem (11.14)! Therefore $\mathfrak{R}_0$ *necessarily* consists of equal and opposite non-zero *pairs* of numbers $\pm\alpha, \pm\beta, \ldots$; and in the α^{th} row of (g_{jk}) the only non-zero element occurs for $k = -\alpha$:

$$g_{\alpha,-\alpha} = -\sum_{a=1}^{l} \alpha_a c^a_{-\alpha,\alpha} + \sum_{\substack{\beta\in\mathfrak{R}^0 \\ (\alpha+\beta\in\mathfrak{R}^0)}} N_{\alpha\beta}N_{-\alpha,\alpha+\beta} + \sum_{a=1}^{l} c^a_{\alpha,-\alpha}c^{-\alpha}_{-\alpha,a} \tag{11.33}$$

The evaluation of the last term here involves some subtlety. It is necessary to *prove* that the l-component vector $(-\alpha)_a$ in $\mathfrak{R}$, associated with $-\alpha \in \mathfrak{R}_0$ is indeed $-\alpha_a$. To see this we start with Eq.(11.30) and apply the Jacobi identity for H_a, E_α and $E_{-\alpha}$.

$$\begin{aligned} 0 &= [H_a, [E_\alpha, E_{-\alpha}]] \\ &= [E_\alpha, [H_a, E_{-\alpha}]] - [E_{-\alpha}, [H_a, E_\alpha]] \\ &= (-\alpha)_a [E_\alpha, E_{-\alpha}] + \alpha_a [E_\alpha, E_{-\alpha}] \\ &= (\alpha_a + (-\alpha)_a)\alpha^b H_b \end{aligned} \tag{11.34}$$

We will show (in point seven to follow) that $\{\alpha^a\}$ cannot vanish identically, so we can conclude:

$$c^{-\alpha}_{a,-\alpha} = (-\alpha)_a = -\alpha_a \quad \text{no sum on } \alpha \tag{11.35}$$

Using this in Eq.(11.33), we see that

$$g_{\alpha,-\alpha} = 2\sum_a \alpha_a \alpha^a - \sum_{\substack{\beta\in\mathfrak{R}_0 \\ (\alpha+\beta\in\mathfrak{R}_0)}} N_{\alpha\beta}N_{\alpha+\beta,-\alpha} \neq 0 \tag{11.36}$$

and the nonsingular $r \times r$ matrix (g_{jk}) has a block diagonal form:

$$(g_{jk}) = \begin{pmatrix} g_{ab} & \cdot & 0 & & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & \cdot & ( ) & 0 & \\ 0 & \cdot & & ( ) & \\ & \cdot & 0 & & \ddots \end{pmatrix},$$

$$\pm\,\alpha^{\text{th}}\ 2\times 2 \text{ block in lower right hand block} = \begin{pmatrix} 0 & g_{\alpha,-\alpha} \\ g_{\alpha,-\alpha} & 0 \end{pmatrix} \tag{11.37}$$

It must then be the case that

$$\det(g_{ab}) \neq 0,$$
$$r - l = \text{even integer} \tag{11.38}$$

We can get an expression for the matrix elements g_{ab}:

$$\begin{aligned} g_{ab} &= \sum_{j,k} c^j_{ak} c^k_{bj} \\ &= \sum_j \sum_{\alpha \in \mathfrak{R}_0} c^j_{a\alpha} c^\alpha_{bj} \\ &= \sum_{\alpha \in \mathfrak{R}_0} c^\alpha_{a\alpha} c^\alpha_{b\alpha} \\ &= \sum_{\alpha \in \mathfrak{R}_0} \alpha_a \alpha_b \end{aligned} \tag{11.39}$$

Since (g_{ab}) is nonsingular, this means that there are enough independent vectors among the $(r-l)$vectors $\{\alpha_a\} \in \mathfrak{R}$ to span l-dimensional space; further, since these vectors come in pairs $\pm\underline{\alpha}$, it must be that

$$\frac{1}{2}(r-l) \geq l,$$

$$\text{i.e.,} \qquad \frac{1}{2}(r-3l) = \text{ integer } \geq 0 \tag{11.40}$$

7. Let us relate the l-component quantities α^a occurring in Eq.(11.30) to the earlier α_a; the former are connected with $[E_\alpha, E_{-\alpha}]$, the latter with $[H_a, E_\alpha]$. We exploit the total antisymmetry of c_{jkm} for this purpose:

$$\begin{aligned} c_{\alpha,-\alpha,a} &= -c_{\alpha,a,-\alpha} \\ \text{i.e.,} \quad c^j_{\alpha,-\alpha} g_{ja} &= -c^j_{\alpha,a} g_{j,-\alpha} \\ \text{i.e.,} \quad g_{ab}\alpha^b &= g_{\alpha,-\alpha}\alpha_a \end{aligned} \tag{11.41}$$

Therefore, as already mentioned in point six above,

$$|g_{ab}| \neq 0, g_{\alpha,-\alpha} \neq 0, \{\alpha_a\} \neq 0 \Rightarrow \{\alpha^a\} \neq 0 \tag{11.42}$$

8. Keeping the H_a unchanged, what happens if we exploit the freedom to change the scale of each E_α independently? Under the change,

$$E_\alpha \to E'_\alpha = n_\alpha E_\alpha, \quad \text{no sum on } \alpha \tag{11.43}$$

we clearly get:

$$\begin{aligned} \alpha'_a &= \alpha_a, \\ g'_{ab} &= g_{ab}, \\ \alpha'^a &= n_\alpha n_{-\alpha} \alpha^a, \\ g'_{\alpha,-\alpha} &= n_\alpha n_{-\alpha} g_{\alpha,-\alpha}, \\ N'_{\alpha\beta} &= n_\alpha n_\beta N_{\alpha\beta}/n_{\alpha+\beta} \end{aligned} \tag{11.44}$$

If we therefore choose the numerical factors n_α to achieve $g_{\alpha,-\alpha} = 1$, then we get:

$$\begin{aligned} \alpha_a &= g_{ab}\alpha^b, \\ \alpha^a &= g^{ab}\alpha_b, \\ (g^{ab}) &= (g_{ab})^{-1}, \\ 2\sum_a \alpha_a\alpha^a - \sum_{\substack{\beta\in\mathfrak{R}_0 \\ (\alpha+\beta\in\mathfrak{R}_0)}} N_{\alpha\beta}N_{\alpha+\beta,-\alpha} &= 1 \end{aligned} \tag{11.45}$$

9. Some useful information on the $N_{\alpha\beta}$ be obtained from the Jacobi identities. Suppose $\alpha, \beta, \alpha+\beta \in \mathfrak{R}_0$, and they are all distinct; then the Jacobi identity for $E_\alpha, E_\beta, E_{-\alpha-\beta}$ and the fact that the vectors $\underline{\alpha}, \underline{\beta}$, are independent gives:

$$N_{\alpha\beta} = N_{\beta,-\alpha-\beta} = -N_{\alpha,-\alpha-\beta} \tag{11.46}$$

10. We can now put together all the information we have gathered on the structure of the Lie Brackets in a complex simple Lie algebra $\tilde{\mathcal{L}}$. The algebra is spanned by $H_a, a = 1, 2, \ldots, l$ and $E_\alpha, \alpha \in \mathfrak{R}_0$, and the brackets among them are:

$$\begin{aligned} [H_a, H_b] &= 0, \\ [H_a, E_\alpha] &= \alpha_a E_\alpha, \\ [E_\alpha, E_{-\alpha}] &= \alpha^a H_a, \\ [E_\alpha, E_\beta] &= N_{\alpha,\beta}E_{\alpha+\beta}, \alpha, \beta, \alpha+\beta \in \mathfrak{R}_0 \end{aligned} \tag{11.47}$$

These equations are called the *Cartan–Weyl* form for any complex simple Lie algebra.

One can proceed in this way to get more information on the roots $\underline{\alpha}, \underline{\beta}, \ldots \in \mathfrak{R}$, their geometrical properties, the structure constants $N_{\alpha\beta}$ etc. For instance, if $\underline{\alpha}, \underline{\beta} \in \mathfrak{R}$, one can ask under what conditions $\underline{\beta} \pm \underline{\alpha}, \underline{\beta} \pm 2\underline{\alpha}, \ldots$ belong to $\mathfrak{R}$ as well. Such an analysis is presented for example by Racah. At this point, however, we will switch attention to the real simple compact Lie algebras.

11.3 The Real Compact Simple Lie Algebras

To descend from Cartan's classification of complex simple Lie algebras to the real compact simple ones, we need to depend on some fundamental theorems:

I: Every complex simple Lie algebra $\tilde{\mathcal{L}}$ has bases in which all structure constants become real. Thus $\tilde{\mathcal{L}}$ definitely can be obtained by a process of complexification of (several) real Lie algebras $\mathcal{L}, \mathcal{L}', \ldots$.

II. (Weyl): Every complex simple Lie algebra $\tilde{\mathcal{L}}$ has a *unique* compact simple real form $\mathcal{L}$. Compactness here means that the Lie groups G associated with $\mathcal{L}$ are compact.

III (Weyl): If G is a (real) compact simple Lie group, and G' is locally isomorphic to G, then G' is also compact. Thus compactness of a real simple Lie group can be read off from its Lie algebra, which further illuminates Theorem II above.

We will now state various properties of (real) compact simple Lie algebras, which taken together with the results already obtained in the complex case, will permit a complete classification of all possible CSLA's. In the following, $\mathcal{L}$ will denote some real compact simple Lie algebra of dimension r and rank l.

1. There exist bases for $\mathcal{L}$ (real, of course) in which the (real) structure constants c_{jk}^m become *totally antisymmetric* and can be written as c_{jkm}. This is because the real metric tensor g_{jk} is positive definite and we can choose the basis so that it becomes the unit tensor:

$$\begin{aligned} g_{jk} &= \delta_{jk} \\ c_{jk}^m &= c_{jkm} \end{aligned} \tag{11.48}$$

Therefore the adjoint representation generators X_j are i times real antisymmetric matrices, so at the level of the group the adjoint representation matrices are real orthogonal. Incidentally the adjoint representation is irreducible because $\mathcal{L}$ is simple.

2. All matrix representations of $\mathcal{L}$ are by hermitian matrices; in particular, any irreducible representation is, with no loss of generality, by finite dimensional hermitian matrices. These statements of hermiticity refer of course to the real elements of $\mathcal{L}$. The Cartan–Weyl basis however has the following behaviour:

$$H_a^\dagger = H_a, \quad E_\alpha^\dagger = E_{-\alpha} \tag{11.49}$$

Thus, while the H_a are indeed real elements of $\mathcal{L}$, the E_α are complex combinations of real elements, and in that sense, strictly speaking, they are not elements of $\mathcal{L}$ at all. They are analogous to the raising and lowering operators in angular momentum theory. Nevertheless we will use them in our analysis.

3. Since the H_a are hermitian and mutually commute, in any UIR (to save on words we use the same description as for a group) they can all be simultaneously diagonalised. They can be taken as part of a complete commuting set. The H_a are a maximal commuting or abelian subalgebra of $\mathcal{L}$; it is called a *Cartan subalgebra.*

4. The quantities α_a, α^a and $N_{\alpha\beta}$ are all real. Now from the expression (11.39) for g_{ab}, we can see that by subjecting the H_a to a real orthogonal rotation in l dimensions and then rescaling them, we can arrange for g_{ab} to become δ_{ab}. We can also reduce each $g_{\alpha,-\alpha}$ to unity. When these have been done, the Cartan–Weyl form of the Lie brackets is

$$\begin{aligned} [H_a, H_b] &= 0, \\ [H_a, E_\alpha] &= \alpha_a E_\alpha, \\ [E_\alpha, E_{-\alpha}] &= \alpha_a H_a, \\ [E_\alpha, E_\beta] &= N_{\alpha\beta} E_{\alpha+\beta} \end{aligned} \tag{11.50}$$

We will prove later that (Eq.(12.15))

$$\alpha, \beta, \alpha + \beta \in \mathfrak{R}_0 \Rightarrow N_{\alpha\beta} \neq 0 \tag{11.51}$$

5. The hermiticity property (11.49) for the complex quantities E_α has as a consequence the relation

$$N_{\alpha,\beta} = N_{-\beta,-\alpha} \tag{11.52}$$

Once we have arranged $g_{ab} = \delta_{ab}, g_{\alpha,-\alpha} = 1$, the relation (11.45) can then be simplified to read:

$$\begin{aligned}
\alpha \in \mathfrak{R}_0 : & 2\sum_{a=1}^{l}(\alpha_a)^2 + \sum_{\substack{\beta\in\mathfrak{R}_0 \\ (\alpha+\beta\in\mathfrak{R}_0)}} N_{\alpha,\beta}N_{-\alpha,\alpha+\beta} \\
& = 2\sum_{a=1}^{l}(\alpha_a)^2 + \sum_{\substack{\beta\in\mathfrak{R}_0 \\ (\alpha+\beta\in\mathfrak{R}_0)}} N_{\alpha,\beta}N_{-\alpha-\beta,\alpha} \quad \text{by Eq.(11.52)} \\
& = 2\sum_{a=1}^{l}(\alpha_a)^2 + \sum_{\substack{\beta\in\mathfrak{R}_0 \\ (\alpha+\beta\in\mathfrak{R}_0)}} N^2_{\alpha,\beta} \quad \text{by Eq.(11.46)} \\
& = 1
\end{aligned} \tag{11.53}$$

We have now assembled enough factual information about the CSLA's to permit their complete analysis and classification. This will be begun in the next Chapter.

Exercises for Chapter 11

1. The Lie algebras of SO(3) and of the three dimensional homogeneous Lorentz group SO(2, 1) are respectively

$$[J_1, J_2] = iJ_3, [J_2, J_3] = iJ_1, [J_3, J_1] = iJ_2;$$

$$[J_0, K_1] = iK_2, [J_0, K_2] = -iK_1, [K_1, K_2] = -iJ_0;$$

Show that they have a common complex extension.

2. For the two groups SO(3) and SO(2, 1) whose Lie algebras are given in problem (1) above, find $r, l, R_0, \underline{\alpha}, N_{\alpha\beta}$ of Eq.(11.47). Assume $H = J_3$ or J_0 respectively.

3. Supply the proofs of Eqs.(11.25)-(11.28).

Chapter 12

Geometry of Roots for Compact Simple Lie Algebras

We have seen that in any UIR of a compact simple Lie algebra $\mathcal{L}$ the hermitian Cartan subalgebra generators H_a can be simultaneously diagonalised. A set of simultaneous eigenvalues for the H_a can be written as an l-component real vector $\mu = \{\mu_a\}$. Such vectors are called *weights*. We shall explore their properties in Chapter 16. But it must be clear that root vectors are the weight vectors appearing in the adjoint representation; and by one of the Cartan theorems, the non-zero roots are nondegenerate.

Having scaled and arranged our generators so that

$$g_{ab} = \sum_{\alpha \in \mathfrak{R}_0} \alpha_a \alpha_b = \mathrm{Tr} H_a^{(\mathrm{adj})} H_b^{(\mathrm{adj})} = \delta_{ab}, \tag{12.1}$$

we can say that the roots of a compact simple Lie algebra of rank l are real vectors in an l-dimensional real Euclidean space. The set $\mathfrak{R}$ contains $r - l$ non-zero distinct root vectors, coming in pairs $\pm\underline{\alpha}$. As these vectors must obey Eq.(12.1), there must be exactly l linearly independent vectors in $\mathfrak{R}$ so that, as we saw earlier in Eq.(11.40), $\frac{1}{2}(r - 3l)$ must be a non-negative integer.

What arrays or geometrical arrangements of distinct pairs of vectors $\pm\underline{\alpha}$ in l-dimensional Euclidean space can arise in the set of roots $\mathfrak{R}$ of some CSLA of rank l? It turns out that there are many restrictions on such arrays, again on account of the Jacobi conditions. A particularly important part is played by a family of (in general not mutually commuting) SU(2) subalgebras, so let us begin by recording the structure of the SU(2) Lie algebra and its (hermitian) irreducible representations.

In the Cartan–Weyl form, the elements of the SU(2) algebra are J_+, J_- and

J_3 obeying the following hermiticity and commutation properties:

$$\begin{aligned} J_3^\dagger = J_3, J_+^\dagger = J_-, \\ [J_3, J_\pm] = \pm J_\pm, \\ \text{and} \quad [J_+, J_-] = 2J_3 \end{aligned} \tag{12.2}$$

The UIR's are labelled by a quantum number j which can take values $0,\ 1/2,\ 1,\ 3/2,\ \ldots$. The j^{th} UIR is of dimension $2j+1$. Within this UIR, the spectrum of J_3 is nondegenerate and consists of the string of eigenvalues $m = j, j-1, j-2, \ldots, -j+1, -j$:

$$J_3|j,m>=m|j,m>; \tag{12.3}$$

zero occurs if j is integral and does not if it is half odd integral. The action of $J_\pm$ is

$$J_\pm|j,m>=\sqrt{(j\mp m)(j\pm m+1)}|j,m\pm 1> \tag{12.4}$$

That is, while each $|j,m>$ is normalised to unity, the phases can be chosen so that $J_\pm$ act in this way.

Let us now identify the different SU(2) algebras contained in a CSLA $\mathcal{L}$. There are in fact $\frac{1}{2}(r-l)$ distinct ones. For each root vector $\underline{\alpha} \in \mathfrak{R}$, we have the subset of commutation relations,

$$\begin{aligned} [H_a, E_\alpha] = \alpha_a E_\alpha, \\ [H_a, E_{-\alpha}] = -\alpha_a E_{-\alpha}, \\ \text{and} \qquad [E_\alpha, E_{-\alpha}] = \alpha_a H_a \end{aligned} \tag{12.5}$$

We can see then that for each $\alpha \in \mathfrak{R}_0$ there is an $\mathrm{SU}(2)^{(\alpha)}$ algebra:

$$\mathrm{SU}(2)^{(\alpha)} : J_3 \to \frac{\alpha_a H_a}{|\underline{\alpha}|^2}, J_\pm \to \frac{\sqrt{2}}{|\underline{\alpha}|} E_{\pm\alpha} \tag{12.6}$$

Further, the hermiticity relations for these $\mathrm{SU}(2)^{(\alpha)}$ generators are also in accordance with Eq.(12.2). The relation between $\mathrm{SU}(2)^{(\alpha)}$ and $\mathrm{SU}(2)^{(-\alpha)}$ is very simple:

$$\mathrm{SU}(2)^{(\alpha)} \to \mathrm{SU}(2)^{(-\alpha)} : J_3 \to -J_3, J_+ \to J_-, J_- \to J_+ \tag{12.7}$$

This is the same as a rotation by π about the x-axis, in the usual language of quantum mechanics and angular momentum.

In the sequel, we choose some $\alpha \in \mathfrak{R}_0, \underline{\alpha} \in \mathfrak{R}$ and keep it fixed. For this reason we have not put α as an index on $J_3, J_\pm$ in Eq.(12.6). We are interested in the action of $\mathrm{SU}(2)^{(\alpha)}$ in the adjoint representation of $\mathcal{L}$: action by J_+, J_- or J_3 on, say, some E_β is the same as commutation of J_+, J_- or J_3 with E_β. Commuting $J_\pm$ with E_β should give as a result $E_{\beta\pm\alpha}$ provided $\beta \pm \alpha \in \mathfrak{R}_0$. We must also remember that for each $\beta \in \mathfrak{R}_0, E_\beta$ is unique apart from a scale, and

that the most general irreducible representation of SU(2) is characterised by a j-value and has the form in Eqs.(12.3),(12.4).

Since all this is so, we can conclude: for any $\underline{\beta} \in \mathfrak{R}, E_\beta$ belongs to some definite UIR of $SU(2)^{(\alpha)}$, carrying definite values of j and m. That is, with respect to commutation with the $SU(2)^{(\alpha)}$ generators, E_β is the m^{th} component of a tensor operator of rank j. (However the relative scales of the different components may not yet be in accord with the matrix element values in Eq.(12.4)). There must be a string of roots, with $p+q+1$ entries, such that

$$\begin{aligned} &\underline{\beta}+p\underline{\alpha}, \underline{\beta}+(p-1)\underline{\alpha}, \ldots, \underline{\beta}+\underline{\alpha}, \underline{\beta}, \underline{\beta}-\underline{\alpha}, \ldots, \underline{\beta}-(q-1)\underline{\alpha}, \\ &\underline{\beta}-q\underline{\alpha} \in \mathfrak{R}, \underline{\beta}+(p+1)\underline{\alpha}, \underline{\beta}-(q+1)\underline{\alpha} \notin \mathfrak{R} \end{aligned} \tag{12.8}$$

The point is that we are using our knowledge of all possible SU(2) representations, and the nondegeneracy of each $\underline{\beta} \in \mathfrak{R}$, to maximum advantage. Under commutation with $J_\pm, J_3$ the string of (complex) generators,

$$E_{\beta+p\alpha}, E_{\beta+(p-1)\alpha}, \cdots, E_{\beta+\alpha}, E_\beta, E_{\beta-\alpha}, \ldots, E_{\beta-(q-1)\alpha}, E_{\beta-q\alpha} \tag{12.9}$$

must behave as the (not yet properly normalised) components of an $SU(2)^{(\alpha)}$ tensor operator. In both Eqs.(12.8),(12.9), it is necessarily true that $p \geq 0, q \geq 0$ are integers; and for the moment we are assuming that $\underline{0}$ does *not* occur in the set of root vectors (12.8). We can easily read off the m-value for E_β, and the rank j for the chain (12.9):

$$\begin{aligned} &[J_3, E_\beta] = (\underline{\alpha} \cdot \underline{\beta}/|\underline{\alpha}|^2)E_\beta \Rightarrow m = \underline{\alpha} \cdot \underline{\beta}/|\underline{\alpha}|^2; \\ &2j+1 = p+q+1 \Rightarrow j = \frac{1}{2}(q+p), m = \frac{1}{2}(q-p) \end{aligned} \tag{12.10}$$

Incidentally, the general structure of SU(2) representations tells us that there can be no gaps either in the string of root vectors (12.8) or generators (12.9).

In case the vector $\underline{0}$ occurs in the string (12.8), we cannot say that all the vectors in the string belong to $\mathfrak{R}$. But you can easily convince yourself that the corresponding generator would have to be $\underline{\alpha} \cdot \underline{H}/|\underline{\alpha}|^2$, namely J_3. From here we can only move "up or down" one step, to E_α or $E_{-\alpha}$, so that $j = 1$ in this case.

What we have therefore done is to use SU(2) representation theory to analyse the geometry of the root vectors as we move up and down in l-dimensional Euclidean space by amounts $\pm\underline{\alpha}, \pm 2\underline{\alpha}, \pm 3\underline{\alpha}, \ldots$. We find that the triplet $E_{\pm\alpha}, \underline{\alpha} \cdot \underline{H}$ forms an (unnormalised) $j = 1$ operator under $SU(2)^{(\alpha)}$. Every $\underline{\beta} \in \mathfrak{R}$ other than $\pm\underline{\alpha}$ belongs to some chain (12.8) of *non-zero roots*; this does not conflict with $m = 0$ being a possible value of m if j is an integer, since it only means that $\underline{\alpha}$ and $\underline{\beta}$ are perpendicular.

Some information on the $N_{\alpha\beta}$ can be obtained, independent of the relative normalisations of the E_β's. With j and m identified as in Eq.(12.10), we do have

$$[J_-, [J_+, E_\beta]] = (j-m)(j+m+1)E_\beta \tag{12.11}$$

since

$$\begin{aligned} [E_\alpha, E_\beta] &= N_{\alpha\beta} E_{\alpha+\beta}, \\ [E_{-\alpha}, E_{\alpha+\beta}] &= N_{-\alpha,\alpha+\beta} E_\beta, \\ \text{and} \quad N_{-\alpha,\alpha+\beta} &= N_{\alpha,\beta} \end{aligned} \tag{12.12}$$

We can get the magnitude of $N_{\alpha\beta}$:

$$N_{\alpha\beta}^2 = \frac{|\underline{\alpha}|^2}{2} p(q+1) \tag{12.13}$$

So we can definitely say

$$\alpha, \beta, \alpha+\beta \in \mathfrak{R}_0 \Rightarrow p \geq 1 \Rightarrow N_{\alpha\beta} \neq 0 \tag{12.14}$$

This property of $N_{\alpha\beta}$ was quoted earlier in Eq.(11.51).

The fact that m appearing in Eq.(12.10) is quantised is the origin of the severe geometrical restrictions on possible systems of root vectors. Before deriving those restrictions, let us record the following additional properties of roots:

$$\begin{aligned} &\underline{\alpha} \in \mathfrak{R} \Rightarrow \pm 2\underline{\alpha}, \pm 3\underline{\alpha}, \ldots \notin \mathfrak{R}, \\ &\underline{\beta} = \underline{\alpha} \text{ in Eq.(12.10)} \Rightarrow p = 0, q = 2, j = m = 1; \\ &\underline{\beta} = -\underline{\alpha} \text{ in Eq.(12.10)} \Rightarrow p = 2, q = 0, j = -m = 1 \end{aligned} \tag{12.15}$$

Now all that we did to E_β by commutation with the $SU(2)^{(\alpha)}$ generators can be repeated with the roles of α and β interchanged! So for *any* two root vectors $\underline{\alpha}, \underline{\beta} \in \mathfrak{R}$ distinct or the same, $SU(2)^{(\alpha)}$ analysis tells us there must exist integers $p, q \geq 0$; while $SU(2)^{(\beta)}$ analysis tells us there must exist integers $p', q' \geq 0$; and then

$$\begin{aligned} \underline{\alpha} \cdot \underline{\beta}/|\underline{\alpha}|^2 &= \frac{1}{2}(q-p) = \frac{n}{2} = \text{positive or negative half integer or zero;} \\ \underline{\alpha} \cdot \underline{\beta}/|\underline{\beta}|^2 &= \frac{1}{2}(q'-p') = \frac{n'}{2} = \text{positive or negative half integer or zero.} \end{aligned} \tag{12.16}$$

If θ is the angle between $\underline{\alpha}$ and $\underline{\beta}$, these restrictions mean:

$$\begin{aligned} \frac{|\underline{\beta}|}{|\underline{\alpha}|}\cos\theta = \frac{n}{2}, \frac{|\underline{\alpha}|}{|\underline{\beta}|}\cos\theta = \frac{n'}{2}, \\ \cos^2\theta = \frac{nn'}{4} \end{aligned} \tag{12.17}$$

We can draw several conclusions on the possible values of n and n':

$$\begin{aligned} &n = 0 \Leftrightarrow n' = 0 \Leftrightarrow \theta = 90°; \\ &n > 0 \Leftrightarrow n' > 0; \\ &n < 0 \Leftrightarrow n' < 0; \\ &n, n' = 0, \pm 1, \pm 2, \pm 3, \pm 4; \\ &0 \leq nn' \leq 4. \end{aligned} \tag{12.18}$$

In those cases where both n and n' are non-zero,

$$|\underline{\alpha}|/|\underline{\beta}| = \sqrt{n'/n} \tag{12.19}$$

We see that the possible values of angles θ between roots, and ratios of lengths of roots, are *severely limited*. Let us see what are the allowed possibilities.

To start, in principle the pair (n, n') can have 17 possible values:

$$\begin{aligned}(n,n') = (0,0); &\pm(1,1); \pm(1,2); \pm(1,3); \pm(1,4); \\ &\pm(2,1); \pm(2,2); \pm(3,1); \pm(4,1)\end{aligned} \tag{12.20}$$

But some of these are actually ruled out! The inadmissible ones are $(n, n') = \pm(1,4)$ and $\pm(4,1)$. For instance, because of Eq.(12.15),

$$n = 1, n' = 4 \Rightarrow \theta = 0 \Rightarrow \underline{\alpha} = 2\underline{\beta} \notin \mathfrak{R} \tag{12.21}$$

and similarly in the other three cases. That leaves 13 possibilities. Of these,

$$(n,n') = \pm(2,2) \Rightarrow \theta = 0 \text{ or } 180°, \quad \underline{\alpha} = \pm\underline{\beta} \tag{12.22}$$

Leaving these aside as well, we are left with 11 significant possibilities, which we present in tabular form:

Geometrical conditions on roots in a CSLA

n	n'	θ	$\lvert\underline{\alpha}\rvert/\lvert\underline{\beta}\rvert$
0	0	90°	unspecified
1	1	60°	1
−1	−1	120°	1
1	2	45°	$\sqrt{2}$
2	1	45°	$1/\sqrt{2}$
−1	−2	135°	$\sqrt{2}$
−2	−1	135°	$1/\sqrt{2}$
1	3	30°	$\sqrt{3}$
3	1	30°	$1/\sqrt{3}$
−1	−3	150°	$\sqrt{3}$
−3	−1	150°	$1/\sqrt{3}$

To find and classify all possible CSLA means to find all possible sets of real root vectors in Euclidean l-dimensional space, for various ranks l, obeying all these geometrical restrictions. This is made tractable by the concepts of *positive* and *simple* roots, to which we devote the next Chapter.

Chapter 13

Positive Roots, Simple Roots, Dynkin Diagrams

13.1 Positive Roots

Let us make a choice of the Cartan subalgebra generators $H_a, a = 1, 2, \ldots, l$ in a definite sequence (of course maintaining $g_{ab} = \delta_{ab}$), and hereafter keep the sequence unchanged. We shall say that a root $\underline{\alpha} \in \mathfrak{R}$ is *positive* if in its set of components $\underline{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_l\}$, the first non-vanishing entry is positive. Otherwise, we shall say $\underline{\alpha}$ is *negative*. The set of all positive roots will be denoted by $\mathfrak{R}_+$, while the set of negative ones will be denoted $\mathfrak{R}_-$. Remembering that every $\underline{\alpha} \in \mathfrak{R}$ is nonzero, and that all roots come in pairs $\pm\underline{\alpha}$, it is clear that every $\underline{\alpha} \in \mathfrak{R}$ is definitely either positive or negative, and each of $\mathfrak{R}_+$ and $\mathfrak{R}_-$ has precisely $\frac{1}{2}(r - l)$ vectors:

$$\begin{aligned} \mathfrak{R} &= \mathfrak{R}_+ \cup \mathfrak{R}_-, \\ \mathfrak{R} &= \text{set of } r - l \quad \text{distinct root vectors,} \\ \mathfrak{R}_+(\mathfrak{R}_-) &= \text{set of } \frac{1}{2}(r - l) \quad \text{distinct positive (negative) root vectors:} \\ \underline{\alpha} \in \mathfrak{R}_+ &\Leftrightarrow -\underline{\alpha} \in \mathfrak{R}_- \end{aligned} \tag{13.1}$$

13.2 Simple Roots and their Properties

Now we define simple roots: these are a subset of the set $\mathfrak{R}_+$ of positive roots, so the simple roots form a set $\mathcal{S} \subset \mathfrak{R}_+$. A root $\underline{\alpha} \in \mathfrak{R}_+$ is called a *simple root*, $\underline{\alpha} \in \mathcal{S}$, if and only if $\underline{\alpha}$ cannot be expressed as a non-negative integral linear combination of other positive roots. We have the inclusion relations

$$\mathfrak{R} = \text{all roots} \supset \mathfrak{R}_+ = \text{all positive roots} \supset \mathcal{S} = \text{all simple roots.} \tag{13.2}$$

A finer subdivision of $\mathfrak{R}_+$ will appear later on.

The properties of $\mathcal{S}$ are really remarkable and beautiful. We describe and prove them one by one.

(i) The angles between simple roots are further restricted, beyond the restrictions listed in the previous Chapter. Let $\underline{\alpha}, \underline{\beta}$ be two distinct simple roots. Then neither $\underline{\alpha} - \underline{\beta}$ nor $\underline{\beta} - \underline{\alpha}$ can be a root. Otherwise, one of them would be in $\mathfrak{R}_+$ and the other in $\mathfrak{R}_-$. If $\underline{\alpha} - \underline{\beta} \in \mathfrak{R}_+$, we are able to express

$$\underline{\alpha} = \underline{\alpha} - \underline{\beta} + \underline{\beta}, \tag{13.3}$$

so $\underline{\alpha}$ could not have been simple. Similarly, $\underline{\beta} - \underline{\alpha} \in \mathfrak{R}_+$ would have led to the conclusion that $\underline{\beta}$ is not simple. We can state the result as

$$\underline{\alpha}, \underline{\beta} \in \mathcal{S}, \underline{\alpha} \neq \underline{\beta} \Rightarrow \underline{\alpha} - \underline{\beta}, \underline{\beta} - \underline{\alpha} \notin \mathfrak{R} \tag{13.4}$$

Now let us interpret this in the light of the actions of the $SU(2)^{(\alpha)}$ and $SU(2)^{(\beta)}$ algebras on $\mathfrak{R}$. Using the notations of the previous Chapter,

$$SU(2)^{(\alpha)} \text{ applied to } E_\beta : q = 0, p \geq 0;$$

$$SU(2)^{(\beta)} \text{ applied to } E_\alpha : q' = 0, p' \geq 0;$$

$$\frac{\underline{\alpha} \cdot \underline{\beta}}{|\underline{\alpha}|^2} = -\frac{p}{2} \leq 0, \frac{\underline{\alpha} \cdot \underline{\beta}}{|\underline{\beta}|^2} = -\frac{p'}{2} \leq 0 \tag{13.5}$$

Referring to Eqs.(12.16)-(12.19), we have $n = -p$ and $n' = -p'$. So the angle θ between $\underline{\alpha}$ and $\underline{\beta}$ must obey

$$\underline{\alpha}, \underline{\beta} \in \mathcal{S}, \underline{\alpha} \neq \underline{\beta} : 90^\circ \leq \theta < 180^\circ \tag{13.6}$$

(We must rule out $\theta = 180^\circ$, since that means $\underline{\beta} = -\underline{\alpha}$, which conflicts with both being positive roots!) The possible angles and length ratios for simple roots form the following subset of the table in Chapter 12:

Geometrical conditions on simple roots in a CSLA

n	n'	θ	$\lvert\underline{\alpha}\rvert/\lvert\underline{\beta}\rvert$
0	0	90°	unspecified
−1	−1	120°	1
−1	−2	135°	$\sqrt{2}$
−2	−1	135°	$1/\sqrt{2}$
−1	−3	150°	$\sqrt{3}$
−3	−1	150°	$1/\sqrt{3}$

(ii) The set of simple roots is linearly independent. For, suppose there is a nontrivial relation

$$\sum_{\underline{\alpha} \in \mathcal{S}} x_\alpha \underline{\alpha} = 0 \tag{13.7}$$

among the simple roots. Since each $\underline{\alpha}$ here is in $\mathfrak{R}_+$, there *must* be some $x_\alpha > 0$ and other $x_\alpha < 0$. (Naturally terms with $x_\alpha = 0$ can be ignored). Write x'_α

for the former, $-y'_\alpha$ for the latter, and splitting the two sets of terms, write Eq.(13.7) as

$$\underline{\gamma} \equiv \sum_{\underline{\alpha} \in \mathcal{S}} x'_\alpha \underline{\alpha} = \sum_{\underline{\alpha} \in \mathcal{S}} y'_\alpha \underline{\alpha} \neq 0 \tag{13.8}$$

Here the simple roots occurring in the two sums are *definitely distinct.* One then has

$$0 < |\underline{\gamma}|^2 = \sum_{\underline{\alpha}, \underline{\beta} \in \mathcal{S}} x'_\alpha y'_\beta \underline{\alpha} \cdot \underline{\beta} \leq 0 \tag{13.9}$$

because each x'_α, y'_β is strictly positive, and from point (i) above $\underline{\alpha} \cdot \underline{\beta}$ is non-positive. Since situation (13.9) is an impossibility, the result is proved. It is interesting to note that the Euclidean geometry for the space of roots was used in the argument, since that is what ensures $\underline{\gamma} \neq 0 \Rightarrow |\underline{\gamma}| > 0$.

(iii) Each positive root $\underline{\alpha} \in \mathfrak{R}_+$ can be written uniquely as a linear combination of simple roots with non-negative integer coefficients. For, let $\underline{\alpha} \in \mathfrak{R}_+$. If $\underline{\alpha} \in \mathcal{S}$, we are done. If $\underline{\alpha} \notin \mathcal{S}$, it can be written as a positive integral combination of other positive roots (definition of $\mathcal{S}$!). Do so. If each term occurring involves a simple root, we are done. If not, the terms involving nonsimple roots can again be written as a positive integral combination of positive roots. This process *must end* with a positive integral combination of *simple* roots because at each stage all coefficients are positive integers, all vectors are positive, so there cannot be an indefinite or unending build-up of terms. The uniqueness of such an expression follows from property (ii) above of linear independence of simple roots. We can summarise:

$$\begin{aligned} \underline{\alpha} \in \mathfrak{R}_+ &\Rightarrow \underline{\alpha} = \text{unique non-negative integer linear combination of} \\ &\qquad \text{simple roots,} \\ \underline{\alpha} \in \mathfrak{R}_- &\Rightarrow \underline{\alpha} = \text{unique non-positive integer linear combination of} \\ &\qquad \text{simple roots} \end{aligned}$$

(iv) If we now combine properties (ii) and (iii) above with the fact, noted at Eq.(12.1), that there are certainly enough independent vectors in $\mathfrak{R}$ to span l-dimensional Euclidean space, we arrive at the nice result that there are exactly l simple roots in $\mathcal{S}$! Hereafter we write

$$\mathcal{S} = \{\underline{\alpha}^{(1)}, \underline{\alpha}^{(2)}, \ldots, \underline{\alpha}^{(l)}\} = \{\underline{\alpha}^{(a)}\}, a = 1, 2, \ldots, l, \tag{13.10}$$

and the contents of point (iii) above are:

$$\begin{aligned} &\underline{\alpha} \in \mathfrak{R}_+ \Rightarrow \underline{\alpha} = \sum_{a=1}^{l} n_a \underline{\alpha}^{(a)}, n_a \geq 0, \text{ integer, unique for } \underline{\alpha}; \\ &\underline{\alpha} \in \mathfrak{R}_- \Rightarrow \underline{\alpha} = \sum_{a=1}^{l} n_a \underline{\alpha}^{(a)}, n_a \leq 0, \text{ integer, unique for } \underline{\alpha}. \end{aligned} \tag{13.11}$$

The point to notice is that in this way of expressing every root $\underline{\alpha} \in \mathfrak{R}$ in terms of simple roots, there is never any need to deal with expressions with some positive and some negative coefficients!

(v) Once the set $\mathcal{S}$ of simple roots is known, their length-angle relations already determine completely how to build up $\mathfrak{R}_+$, and so then also $\mathfrak{R}_-$. From (iv), each positive root is a unique non-negative integral linear combination of simple roots, so they can be ordered according to the "number of terms". Introducing

$$N = \sum_{a=1}^{l} n_a, \tag{13.12}$$

we define unique subsets of $\mathfrak{R}_+$ thus:

$$\begin{aligned} N = 1 &: \ \mathcal{S} = \text{ subset of simple roots;} \\ N = 2 &: \ \mathfrak{R}_+^{(2)} = \text{ positive roots with } n_1 + n_2 + \ldots + n_l = 2; \\ N = 3 &: \ \mathfrak{R}_+^{(3)} = \text{ positive roots with } n_1 + n_2 + \ldots + n_l = 3; \end{aligned} \tag{13.13}$$

and so on. Thus, $\mathfrak{R}_+$ can be broken up unambiguously as

$$\mathfrak{R}_+ = \mathcal{S} \cup \mathfrak{R}_+^{(2)} \cup \mathfrak{R}_+^{(3)} \ldots \cup \mathfrak{R}_+^{(N)} \cup \mathfrak{R}_+^{(N+1)} \ldots \tag{13.14}$$

(Of course for any given CSLA, there is only a finite number of terms here). We will now show how to build up $\mathfrak{R}_+^{(2)}$ from $\mathcal{S}$, and then by induction $\mathfrak{R}_+^{(N+1)}$ out of $\mathcal{S}, \mathfrak{R}_+^{(2)}, \ldots, \mathfrak{R}_+^{(N)}$.

In Chapter 12, we introduced the set of $\frac{1}{2}(r-l)$ distinct $\mathrm{SU}(2)^{(\alpha)}$ subalgebras in $\mathcal{L}$, one for each pair $\pm\underline{\alpha}$ of roots in $\mathfrak{R}$. Now the simple roots $\{\underline{\alpha}^{(a)}\}$ in $\mathcal{S}$ have been singled out, as having a special significance. We therefore introduce a special notation for these l SU(2) subalgebras:

$$\mathrm{SU}(2)^{(a)} = \ \mathrm{SU}(2)^{(\alpha)} \text{ for } \underline{\alpha} = \underline{\alpha}^{(a)}, a = 1, 2, \ldots, l \tag{13.15}$$

Consider now the construction of $\mathfrak{R}_+^{(2)}$. In an expression of the form (13.11), we cannot have any $n_a = 2$ as the result (12.15) forbids this. So each $\underline{\alpha} \in \mathfrak{R}_+^{(2)}$ must be the sum of two *distinct* simple roots. In general,

$$\underline{\alpha}^{(a)}, \underline{\alpha}^{(b)} \in \mathcal{S}, a \neq b, \underline{\beta} = \underline{\alpha}^{(a)} + \underline{\alpha}^{(b)} \in \mathfrak{R}_+ \tag{13.16}$$

implies (in the same notation as in Eqs.(12.8),(12.16)),

$$\begin{aligned} &\mathrm{SU}(2)^{(a)} \text{ applied to } E_{\alpha^{(b)}} : q = 0, p \geq 1, \frac{\underline{\alpha}^{(a)} \cdot \underline{\alpha}^{(b)}}{|\underline{\alpha}^{(a)}|^2} = -\frac{p}{2}; \\ &\mathrm{SU}(2)^{(b)} \text{ applied to } E_{\alpha^{(a)}} : q' = 0, p' \geq 1, \frac{\underline{\alpha}^{(a)} \cdot \underline{\alpha}^{(b)}}{|\underline{\alpha}^{(b)}|^2} = -\frac{p'}{2} \end{aligned} \tag{13.17}$$

The vanishing of q and q' already follows from Eq.(13.5). We therefore have the following simple rule to determine which pairs of distinct simple roots in $\mathcal{S}$ can be added to produce positive roots in $\mathfrak{R}_+^{(2)}$:

$$\begin{aligned} &\underline{\alpha}^{(a)}, \underline{\alpha}^{(b)} \in \mathcal{S},\ a \neq b: \\ &\quad \underline{\alpha}^{(a)} \cdot \underline{\alpha}^{(b)} = 0 \ \Rightarrow\ \underline{\alpha}^{(a)} + \underline{\alpha}^{(b)} \notin \mathfrak{R}_+^{(2)}, \\ &\quad \underline{\alpha}^{(a)} \cdot \underline{\alpha}^{(b)} < 0 \ \Rightarrow\ \underline{\alpha}^{(a)} + \underline{\alpha}^{(b)} \in \mathfrak{R}_+^{(2)} \end{aligned} \tag{13.18}$$

Any two simple roots making an angle of 120°, 135° or 150° can be added to give a root in $\mathfrak{R}_+^{(2)}$.

Now we use the method of induction. Suppose $\mathfrak{R}_+^{(2)}, \mathfrak{R}_+^{(3)}, \ldots, \mathfrak{R}_+^{(N)}$ have been built up. How do we construct $\mathfrak{R}_+^{(N+1)}$? We want to answer the question: if $\underline{\beta} \in \mathfrak{R}_+^{(N)}$ and $\underline{\alpha}^{(a)} \in \mathcal{S}$, when is $\underline{\beta} + \underline{\alpha}^{(a)} \in \mathfrak{R}_+^{(N+1)}$? We examine the way $\underline{\beta}$ has been built up and ask: how often can $\underline{\alpha}^{(a)}$ be subtracted from $\underline{\beta}$, leaving a positive root as result? We seek the value of q such that:

$$\begin{aligned} \underline{\beta} - \underline{\alpha}^{(a)} \in \mathfrak{R}_+^{(N-1)}, \underline{\beta} - 2\underline{\alpha}^{(a)} \in \mathfrak{R}_+^{(N-2)}, \ldots, \underline{\beta} - q\underline{\alpha}^{(a)} \in \mathfrak{R}_+^{(N-q)}, \\ \underline{\beta} - (q+1)\underline{\alpha}^{(a)} \notin \mathfrak{R}_+ \end{aligned} \tag{13.19}$$

Evidently, we are examining the behaviour of E_β under $\mathrm{SU}(2)^{(a)}$. Because of the uniqueness statements in point (iv) above, especially Eq.(13.11) and the comment following, *as also the fact that we are concerned only with the actions of $SU(2)^{(a)}$ and not of $SU(2)^{(\alpha)}$ for all $\alpha \in \mathfrak{R}_0$*, we can be sure that we need not "descend below $\mathcal{S}$" in the sequence (13.19). There will definitely be a value of q obeying $0 \leq q \leq N-1$ and satisfying (13.19). The value of the index p associated with this $\mathrm{SU}(2)^{(a)}$ action on E_β is of course then fixed, since

$$\frac{\underline{\alpha}^{(a)} \cdot \underline{\beta}}{|\underline{\alpha}^{(a)}|^2} = \frac{1}{2}(q-p) \tag{13.20}$$

If p found in this way is greater than zero, then we can add $\underline{\alpha}^{(a)}$ to $\underline{\beta}$ and get a positive root in $\mathfrak{R}_+^{(N+1)}$, otherwise not:

$$\begin{aligned} p = 0 &: \underline{\beta} + \underline{\alpha}^{(a)} \notin \mathfrak{R}_+^{(N+1)} \\ p > 0 &: \underline{\beta} + \underline{\alpha}^{(a)} \in \mathfrak{R}_+^{(N+1)} \end{aligned} \tag{13.21}$$

Thus, the knowledge of the set $\mathcal{S}$ of simple roots already contains complete information on $\mathfrak{R}_+$, hence also $\mathfrak{R}_-$ and $\mathfrak{R}$.

The simplest nontrivial example of all this is SU(3), for which $r = 8, l = 2$. (The case of SU(2) is left as an exercise to the reader, here $r = 3, l = 1, \mathcal{S} = \mathfrak{R}_+$). We will define and describe the unitary unimodular groups $\mathrm{SU}(l+1)$ in some detail in the next Chapter, but there is no harm in exhibiting the SU(3) case as an illustration of the reconstruction process $\mathcal{S} \to \mathfrak{R}_+^{(2)} \to \mathfrak{R}_+^{(3)} \ldots$. Since $l = 2$,

the root vectors can all be drawn in a plane (and that is recommended as an exercise as well). With suitable normalisations of H_1 and H_2, it turns out that:

$$\text{SU(3): } \mathfrak{R}_+ = \left\{(1,0), \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right), \left(\frac{1}{2}, \frac{-\sqrt{3}}{2}\right)\right\}$$
$$\mathcal{S} = \left\{\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right), \left(\frac{1}{2}, \frac{-\sqrt{3}}{2}\right)\right\} = \{\underline{\alpha}^{(1)}, \underline{\alpha}^{(2)}\}, \text{ say.} \qquad (13.22)$$

The angle between $\underline{\alpha}^{(1)}$ and $\underline{\alpha}^{(2)}$ is 120°, so their sum is an allowed positive root, giving us back $(1,0) \in \mathfrak{R}_+$. Writing $\underline{\beta} = \underline{\alpha}^{(1)} + \underline{\alpha}^{(2)}$, we apply $\text{SU(2)}^{(1)}$ and $\text{SU(2)}^{(2)}$ in turn to E_β:
$\text{SU(2)}^{(1)}$ applied to $E_\beta : q = 1$;

$$\frac{\underline{\alpha}^{(1)} \cdot \underline{\beta}}{|\underline{\alpha}^{(1)}|^2} = \frac{1}{2}(q-p) = \frac{1}{2} \Rightarrow p = 0;$$

$\text{SU(2)}^{(2)}$ applied to $E_\beta : q = 1$

$$\frac{\underline{\alpha}^{(2)} \cdot \underline{\beta}}{|\underline{\alpha}^{(2)}|^2} = \frac{1}{2}(q-p) = \frac{1}{2} \Rightarrow p = 0; \qquad (13.23)$$

Thus, neither $\underline{\alpha}^{(1)}$ nor $\underline{\alpha}^{(2)}$ can be added to $\underline{\beta}$ to produce another positive root in $\mathfrak{R}_+^{(3)}$! All the higher subsets $\mathfrak{R}_+^{(3)}, \mathfrak{R}_+^{(4)}, \ldots$, are empty and the reconstruction concludes at $\mathfrak{R}_+^{(2)}$:

$$\mathfrak{R}_+ = \{\underline{\alpha}^{(1)}, \underline{\alpha}^{(2)}, \underline{\alpha}^{(1)} + \underline{\alpha}^{(2)}\} \qquad (13.24)$$

After we have derived the root systems for the other nontrivial rank 2 CSLA's, the reader can come back and satisfy himself/herself that all this works.

13.3 Dynkin Diagrams

All the geometrical information about the angles and length ratios among the simple roots $\{\underline{\alpha}^{(a)}\} = \mathcal{S}$ of any CSLA $\mathcal{L}$ can be given in a two-dimensional diagram called the *Dynkin diagram*. Because of the reconstruction theorem just established in the previous section, we can in fact say that each possible CSLA $\mathcal{L}$ corresponds to a possible Dynkin diagram, and vice versa. An allowed system of simple roots $\mathcal{S}$ is called a *π-system*, this is the same as an allowed Dynkin diagram.

The rules for constructing a Dynkin diagram are the following: Each simple root $\underline{\alpha}^{(a)} \in \mathcal{S}$ is depicted by a circle ◯. If $\underline{\alpha}^{(a)}$ and $\underline{\alpha}^{(b)}$ are perpendicular to one another, the circles are left unconnected. For an angle $\theta = 120°, 135°, 150°,$

we draw one, two or three lines to connect the respective circles:

$$\begin{aligned}\theta = 90°: &\quad \bigcirc \qquad \bigcirc \\ \theta = 120°: &\quad \bigcirc\!\!-\!\!-\!\!\bigcirc \\ \theta = 135°: &\quad \bigcirc\!\!=\!\!\bigcirc \\ \theta = 150°: &\quad \bigcirc\!\!\equiv\!\!\bigcirc\end{aligned} \tag{13.25}$$

If two circles are doubly or triply connected, the corresponding simple root vectors have lengths in the ratio $\sqrt{2}$ or $\sqrt{3}$ respectively. This can be indicated by, for example, shading the circle for the longer root. We will see examples later.

It is well to remember that a Dynkin diagram depicts some geometrical arrangement of l independent vectors in Euclidean l-dimensional space; and if *any two* circles in the diagram are *unconnected*, those two simple roots are at 90° to one another.

An important fact about Dynkin diagrams, which we now prove, is this: For a *simple* compact Lie algebra $\mathcal{L}$, the Dynkin diagram as a whole must be *connected*, it cannot split into two (or more) disconnected parts. The proof is as follows. Suppose the set of all simple roots, $\mathcal{S}$, splits into the union of two subsets, $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$, with every root in $\mathcal{S}^{(1)}$ perpendicular to every root in $\mathcal{S}^{(2)}$. For the present proof only, let us generically write $\underline{\alpha}$ for simple roots in $\mathcal{S}^{(1)}$, $\underline{\beta}$ for those in $\mathcal{S}^{(2)}$:

$$\begin{aligned}&\mathcal{S} = \mathcal{S}^{(1)} \cup \mathcal{S}^{(2)}; \\ &\underline{\alpha} \in \mathcal{S}^{(1)}, \underline{\beta} \in \mathcal{S}^{(2)} : \underline{\alpha} \cdot \underline{\beta} = 0\end{aligned} \tag{13.26}$$

It follows from Eq.(posrootsinr) that $\mathfrak{R}_+^{(2)}$ consists of vectors like $\underline{\alpha}^{(1)} + \underline{\alpha}^{(2)}, \underline{\beta}^{(1)} + \underline{\beta}^{(2)}$, but *none* of the form $\underline{\alpha} + \underline{\beta}$. Next, in $\mathfrak{R}_+^{(3)}$, can we have a combination like $\underline{\gamma} = \underline{\alpha}^{(1)} + \underline{\alpha}^{(2)} + \underline{\beta}$? If so, it must be that $\underline{\alpha}^{(1)} + \underline{\alpha}^{(2)} \in \mathfrak{R}_+^{(2)}$. Apply $\mathrm{SU}^{(2)(\beta)}$ to this vector in $\mathfrak{R}_+^{(2)}$:
clearly, $q = 0$, *but also* $p = 0$, since

$$(\underline{\alpha}^{(1)} + \underline{\alpha}^{(2)}) \cdot \underline{\beta} = 0 \tag{13.27}$$

So there *could not* have been a vector like $\underline{\gamma}$ above in $\mathfrak{R}_+^{(3)}$! And so on for higher orders. We thus see that if the simple roots $\mathcal{S}$ split into two disjoint sets as in (13.26), so do all possible roots:

$$\begin{aligned}&\mathfrak{R} = \mathfrak{R}^{(1)} \cup \mathfrak{R}^{(2)}, \\ &\mathfrak{R}^{(1)} = \text{ linear combinations of } \underline{\alpha}\text{'s}; \\ &\mathfrak{R}^{(2)} = \text{ linear combinations of } \underline{\beta}\text{'s}; \\ &\underline{\alpha} \in \mathfrak{R}^{(1)}, \underline{\beta} \in \mathfrak{R}^{(2)} \Rightarrow \underline{\alpha} \cdot \underline{\beta} = 0\end{aligned} \tag{13.28}$$

In particular there are *no roots* of the form $\underline{\alpha} + \underline{\beta}$ in $\mathfrak{R}$. Consequently, all the E_α's commute with all the E_β's:

$$\underline{\alpha} \in \mathfrak{R}^{(1)}, \underline{\beta} \in \mathfrak{R}^{(2)} : [E_\alpha, E_\beta] = 0 \tag{13.29}$$

We are close to the final result. We need only to check the properties of the Cartan subalgebra generators H_a. Already we know,

$$\begin{aligned} &\underline{\alpha} \in \mathcal{S}^{(1)}, \underline{\beta} \in \mathfrak{R}^{(2)} : [\underline{\alpha} \cdot \underline{H}, E_\beta] = 0; \\ &\underline{\beta} \in \mathcal{S}^{(2)}, \underline{\alpha} \in \mathfrak{R}^{(1)} : [\underline{\beta} \cdot \underline{H}, E_\alpha] = 0 \end{aligned} \tag{13.30}$$

But now, since the simple roots in $\mathcal{S}$ are independent and span the full space, the H_a can be replaced by the independent combinations $\underline{\alpha} \cdot \underline{H}$ for $\underline{\alpha} \in \mathcal{S}^{(1)}$ and $\underline{\beta} \cdot \underline{H}$ for $\underline{\beta} \in \mathcal{S}^{(2)}$. And then the entire Lie algebra splits into the direct sum of two commuting subalgebras, so it is not simple!

We have outlined the general properties of systems of roots and simple roots, and developed the diagrammatic method which can concisely convey all the essential properties of a simple root system $\mathcal{S}$. The problem of complete classification of all CSLA's is then clear: find all possible "allowed" Dynkin diagrams, i.e., π-systems! This programme will be taken up and completed in Chapter 15. As an interlude, however, we look in the next Chapter at the four classical families of groups SO($2l$), SO($2l+1$), USp($2l$) and SU($l+1$): these are "almost all" of the possible compact simple Lie groups, there being only five others! Apart from providing an interlude, Chapter 14 will introduce us to these groups and their Lie algebras, and give us tangible examples of the general theory of Chapter 12 and the present Chapter.

Exercises for Chapters 12 and 13

1. Supply the proofs of Eqs.(12.13), (12.16)–(12.22).

2. For the common Lie algebra of SO(3) and SU(2), with hermitian generators J_1, J_2, J_3 obeying
$$[J_j, J_k] = i\epsilon_{jkl} J_l,$$
choosing J_3 as the (single) Cartan subalgebra generator H_a with $a = 1$: arrange J_1 and J_2 into suitable complex combinations $E_{\pm\alpha}$, find the set of all roots $R = \{\alpha\}$, the set of positive roots R_+ and of simple roots $\mathcal{S}$. Show that in this case, Eq.(12.22) applies.

Chapter 14

Lie Algebras and Dynkin Diagrams for $\mathbf{SO}(2l)$, $\mathbf{SO}(2l+1)$, $\mathbf{USp}(2l)$, $\mathbf{SU}(l+1)$

Let us begin with some general remarks. For each of the four classical families of groups, we shall start with a defining representation, which is naturally faithful. Throughout this Chapter, we shall uniformly use the symbol D for defining representations. With the matrices of this in hand, we can find and parametrise elements near the identity, so read off the basic Lie bracket relations, identify H_a and $E_\alpha, \mathfrak{R}$ and $\mathfrak{R}_+$ and $\mathcal{S}$ etc. In the defining representation D, as in any UIR, the simultaneous eigenvalue sets for the H_a are the *weights* $\underline{\mu} = \{\mu_a\}$ of that representation. As we will see in more detail in Chapter 16, the general relationship between roots and weights is

$$\begin{aligned} \text{Roots} &\sim \text{ weights of the adjoint representation} \\ &\sim \text{ differences of weights of general representations} \end{aligned} \tag{14.1}$$

This must be kept in mind in what follows.

For each family of groups we will adopt this sequence: defining representation D; infinitesimal generators and commutation relations; Cartan subalgebra generators H_a; weights $\underline{\mu}$ occurring in D; set of all roots $\mathfrak{R}$; positive roots $\mathfrak{R}_+$; simple roots $\mathcal{S}$; the associated Dynkin diagram. Uniformly, the index l denotes the rank of the concerned group.

14.1 The $\mathbf{SO}(2l)$ Family — D_l of Cartan

These are the groups of real, orthogonal, unimodular rotations in Euclidean spaces of even number of dimensions. For $l = 1$, we have rotations in a plane, an abelian group; for $l = 2$, the group SO(4) happens to have a non-simple Lie

algebra. Therefore in discussing the family of CSLA's in the case of SO($2l$) we limit l to $l \geq 3$. The corresponding algebras — more precisely, their complex forms — were called D_l by Cartan.

The matrices S belonging to the defining representation D of SO($2l$) are real, $2l$-dimensional, orthogonal and unimodular:

$$\begin{aligned} S \in \mathrm{SO}(2l)&: S = 2l\text{-dimensional}, \\ S^* &= S \\ S^T S &= 1 \\ \det S &= +1 \end{aligned} \tag{14.2}$$

Each S describes a proper rotation in $2l$-dimensional Euclidean space. Let indices $A, B, C, \ldots$ go over the range $1, 2, \ldots, 2l$. The generator matrices for D can be found by determining the form of an S close to the identity:

$$\begin{aligned} &S \simeq 1 - i\epsilon X, S^T S = 1, S^* = S, |\epsilon| \ll 1 \Rightarrow \\ &X^* = -X, X^T = -X \end{aligned} \tag{14.3}$$

Thus, the most general X is i times a real antisymmetric $2l$-dimensional matrix. We can construct a basis for such matrices quite easily: writing M_{AB} for them, we define

$$(M_{AB})_{CD} = i(\delta_{AC}\delta_{BD} - \delta_{AD}\delta_{BC}), A, B = 1, 2, \ldots, 2l \tag{14.4}$$

Here, only C and D are row and column indices, and in them we have antisymmetry. However, A and B enumerate the various generator matrices, and since we have antisymmetry in them too,

$$M_{AB} = -M_{BA} \tag{14.5}$$

the number of independent generators is $l(2l-1)$. Thus the order of SO($2l$) is $l(2l-1)$. While one might use the M_{AB} for $A < B$, say, as an independent set of generators, it is more symmetrical to use all M_{AB} subject to the conditions (14.5), in general discussions.

The commutation relations among the M_{AB} are:

$$[M_{AB}, M_{CD}] = i(\delta_{BC}M_{AD} - \delta_{AC}M_{BD} + \delta_{BD}M_{CA} - \delta_{AD}M_{CB}) \tag{14.6}$$

These equations completely fix the structure of the Lie algebra D_l — all we need to do is expose its contents suitably! Any UIR of SO($2l$) is generated by *hermitian* M_{AB} (in a suitable complex space of suitable dimension) obeying these same commutations relations.

Towards identifying the elements of the Cartan subalgebra, it is useful to divide the $2l$ values of $A, B, \ldots$ into l pairs: the a^{th} pair consists of the index values $2a-1, 2a$; and a ranges from 1 to l. Within each pair, we can have indices

$r, s, \ldots$ going over just the values 1, 2:

$$\begin{aligned} A, B, \ldots &\to ar, bs, \ldots; \\ a, b, \ldots &= 1, 2, \ldots, l; r, s, \ldots = 1, 2; \\ A &= 2(a-1) + r \end{aligned} \tag{14.7}$$

If we now take

$$\begin{aligned} & H_1 = M_{12}, H_2 = M_{34}, \ldots, H_l = M_{2l-1,2l} \\ \text{i.e.,} \quad & H_a = M_{2a-1,2a} = M_{a1,a2} \end{aligned} \tag{14.8}$$

we do see that because none of the kronecker deltas in Eq.(14.6) "click", these commute with each other:

$$[H_a, H_b] = 0, a, b = 1, 2, \ldots, l \tag{14.9}$$

Geometrically too this is obvious: H_1 generates SO(2) rotations in the 1-2 plane, H_2 in the 3-4 plane, and so on. It can next be easily shown that these H_a *are a maximal commuting subset* of generators. If one takes a general linear combination X of the M_{AB} and imposes the condition that it commute with each of the H_a, one quickly discovers that it must be a linear combination of the H_a's:

$$\begin{aligned} & X = \frac{1}{2} x_{AB} M_{AB}, [X, H_a] = 0, a = 1, 2, \ldots, l \Rightarrow \\ & X = x_{12} H_1 + x_{34} H_2 + \ldots \end{aligned} \tag{14.10}$$

Thus, the H_a do span a Cartan subalgebra of D_l, so the rank is l.

The matrices S of the defining representation of SO(2l) act on 2l-component real Euclidean vectors. To find the weights of this representation D, we must simultaneously diagonalise all the H_a, working in the complex domain if necessary. To clarify the situation, weights in *any* UIR, including the case of D, are l-component real vectors; the corresponding simultaneous eigenvectors of the H_a are vectors in the representation space, and so in the case of D they are 2l-component quantities. Since the H_a are block-diagonal with the forms (in the representation D!):

$$H_1 = i \begin{bmatrix} 0 & 1 & & \\ -1 & 0 & & \\ & & \ddots & \end{bmatrix}, H_2 = i \begin{bmatrix} \ddots & & & \\ & 0 & 1 & \\ & -1 & 0 & \\ & & & \ddots \end{bmatrix}, \ldots, \tag{14.11}$$

it is quite easy to find their simultaneous eigenvalues. If H_1 has eigenvalue ± 1, then $H_2, H_3, \ldots$ have eigenvalues zero; when H_2 has eigenvalue ± 1, $H_1, H_3, \ldots$ have eigenvalues zero; and so on. Let us write $\underline{e}_a, a = 1, 2, \ldots, l$ for the unit

vectors in l-dimensional Euclidean *root and weight* space:

$$\underline{e}_a = (0,0,\ldots,1,0\ldots 0), a = 1,2,\ldots,l$$
$$\uparrow$$
$$a^{\text{th}} \text{ position} \tag{14.12}$$

Then there are $2l$ weights $\{\underline{\mu}\}$ in the representation D, and they are:

$$D \text{ of SO}(2l): \qquad \{\underline{\mu}\} = \{\pm\underline{e}_1, \pm\underline{e}_2, \ldots, \pm\underline{e}_l\} \tag{14.13}$$

Each of these weights is nondegenerate in this representation, and their number correctly gives the dimension of D.

Once we have got the weights in the defining representation D, the general relationship (14.1) between roots and weights suggests that the roots might be of the forms $\pm e_a \pm e_b$, for $a \neq b$, and $\pm 2e_a$. Which of these actually occur in the set $\mathfrak{R}$ of all roots? The number of distinct roots is determined to be $l(2l-1) - l = 2l(l-1)$. We can determine the set $\mathfrak{R}$ as follows. It was already mentioned that an independent set of generators is M_{AB} for $A < B$. Let us list them, using the split index notation $A \to ar, B \to bs$ of Eq.(14.7) as follows:

$$a = b, r = 1, s = 2 : M_{a1,a2} = H_a;$$
$$a < b : M_{a1,b1} = X_{ab}, M_{a1,b2} = Y_{ab}, M_{a2,b1} = Z_{ab}, M_{a2,b2} = W_{ab} \tag{14.14}$$

The subset of commutation relations between H's on the one hand, and X, Y, Z, W on the other, are:

$$\begin{aligned}
[H_a, X_{bc}] &= -i(\delta_{ab} Z_{bc} + \delta_{ac} Y_{bc}), \\
[H_a, Y_{bc}] &= i(-\delta_{ab} W_{bc} + \delta_{ac} X_{bc}), \\
[H_a, Z_{bc}] &= i(\delta_{ab} X_{bc} - \delta_{ac} W_{bc}), \\
[H_a, W_{bc}] &= i(\delta_{ab} Y_{bc} + \delta_{ac} Z_{bc})
\end{aligned} \tag{14.15}$$

Here it is assumed that $b < c$, and *there is no summation on repeated indices on the right hand side.* To find the set $\mathfrak{R}$ of all roots, we must form combinations of X, Y, Z, W which, upon commutation with each H_a, go into multiples of themselves. Some algebra shows that if we define,

$$\begin{aligned}
b < c, \epsilon = \pm 1, \epsilon' = \pm 1; \underline{\alpha} &= \epsilon \underline{e}_b + \epsilon' \underline{e}_c, \\
E_\alpha &= X_{bc} - i\epsilon Z_{bc} - i\epsilon' Y_{bc} - \epsilon\epsilon' W_{bc},
\end{aligned} \tag{14.16}$$

then,

$$[H_a, E_\alpha] = \alpha_a E_\alpha \tag{14.17}$$

By choosing all possible pairs b, c obeying $b < c$; and for each pair all four choices of ϵ, ϵ'; we do get enough combinations from which all the X, Y, Z, W can be recovered. Thus the complete set of roots $\mathfrak{R}$ is:

$$\text{SO}(2l): \qquad \mathfrak{R} = \{\pm\underline{e}_a \pm \underline{e}_b, a < b\} \tag{14.18}$$

We see that the vectors $\pm 2\underline{e}_a$ *do not* appear as roots; and as both a and b run from 1 to l, the number of distinct roots is exactly $2l(l-1)$ as expected.

The subset of *positive* roots is easily identified

$$\mathfrak{R}_+ = \{\underline{e}_a \pm \underline{e}_b, a < b\} \tag{14.19}$$

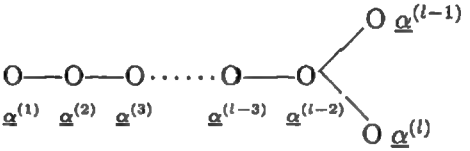
There are $l(l-1)$ of them. To find which of these are *simple*, some analysis is needed. It helps to look at low values of l, and then generalise. One finds:

$$\begin{gathered} \mathcal{S} = \{\underline{e}_1 - \underline{e}_2, \underline{e}_2 - \underline{e}_3, \ldots, \underline{e}_{l-2} - \underline{e}_{l-1}, \underline{e}_{l-1} - \underline{e}_l, \underline{e}_{l-1} + \underline{e}_l\}, \\ \underline{\alpha}^{(1)} = \underline{e}_1 - \underline{e}_2, \underline{\alpha}^{(2)} = \underline{e}_2 - \underline{e}_3, \ldots, \underline{\alpha}^{(l-1)} = \underline{e}_{l-1} - \underline{e}_l, \underline{\alpha}^{(l)} = \underline{e}_{l-1} + \underline{e}_l \end{gathered} \tag{14.20}$$

Each simple root is of length $\sqrt{2}$, so the length ratios are unity. Any two simple roots are either orthogonal or make an angle of 120°; the non-zero scalar products among simple roots are the following:

$$\underline{\alpha}^{(1)} \cdot \underline{\alpha}^{(2)} = \underline{\alpha}^{(2)} \cdot \underline{\alpha}^{(3)} = \ldots = \underline{\alpha}^{(l-2)} \cdot \underline{\alpha}^{(l-1)} = \underline{\alpha}^{(l-2)} \cdot \underline{\alpha}^{(l)} = -1 \tag{14.21}$$

From all this information the Dynkin diagram can be immediately drawn: π-system for SO(2l) ≡ D_l:



(14.22)

Remember that *any two* unconnected circles represent mutually perpendicular simple roots!

14.2 The SO(2l + 1) Family — B_l of Cartan

The preceding analysis of SO(2l) considerably simplifies the work of similarly treating the proper rotation group in an odd number of dimensions, SO(2l + 1). Now the defining representation D consists of (2l + 1)-dimensional real, orthogonal, unimodular matrices:

$$\begin{gathered} S \in \mathrm{SO}(2l+1) : S = (2l+1) \text{ dimensional} \\ S^* = S, \\ S^T S = 1 \\ \det S = +1 \end{gathered} \tag{14.23}$$

These rotations act on (2l + 1)-component Euclidean vectors. Now we let the vector and tensor indices $A, B, \ldots$, go over the range $1, 2, \ldots, 2l+1$: the range appropriate for SO(2l), plus one more value, namely (2l + 1).

By examining the form of an S close to the identity, one finds again that a basis for the Lie algebra in the defining representation consists of matrices $M_{AB} = -M_{BA}$ with the same expression (14.4) for matrix elements; the only difference is that these are now $(2l+1)$-dimensional matrices, and the number of independent matrices is $l(2l+1)$. Thus the order of the group SO$(2l+1)$ is $l(2l+1)$. Even the commutation relations (14.6) retain their validity for the present group, with the ranges of A, B, C, D extended!

One can once again use the split index notation $A \to ar$ to cover the range $1, 2, \ldots, 2l$, and then separately include the value $A = 2l+1$.

The SO$(2l)$ choice of Cartan subalgebra generators $H_a, a = 1, \ldots, l$ serves as a Cartan subalgebra for SO$(2l+1)$ too: even with $(2l+1)$-dimensional matrices, one easily checks that any generator matrix X commuting with $H_1, H_2, \ldots, H_l$ has to be a linear combination of them. Thus SO$(2l+1)$ has rank l. In the present defining representation D, the H_a are the $2l$-dimensional matrices which we had with SO$(2l)$ *plus one extra row and one extra column at the ends consisting entirely of zeros.*. This immediately tells us the weights $\underline{\mu}$ in D: they are the same as with SO$(2l)$, plus the weight $\underline{0}$:

$$D \text{ of SO}(2l+1)\text{: } \{\underline{\mu}\} = \{\pm\underline{e}_1, \pm\underline{e}_2, \ldots, \pm\underline{e}_l, \underline{0}\} \tag{14.24}$$

The number of distinct weights is $2l+1$, so they are all nondegenerate.

Turning to the system of roots $\mathfrak{R}$, since the H_a are "unchanged" in going from SO$(2l)$ to SO$(2l+1)$, all previous roots remain valid; the SO$(2l)$ generators are a subalgebra. The total number of roots is $2l^2$, which exceeds by $2l$ the number of roots for SO$(2l)$. These new roots must arise from the bracket relations between H_a and the extra generators $M_{A,2l+1}$. Using split index notation for the first index here, we find:

$$[H_a, M_{b1,2l+1} - i\epsilon M_{b2,2l+1}] = \epsilon\delta_{ab}(M_{b1,2l+1} - i\epsilon M_{b2,2l+1}),$$
$$\epsilon = \pm 1, \text{ no sum on } b \tag{14.25}$$

Thus we have the new roots $\pm\underline{e}_b, b = 1, 2, \ldots l$ here! So the full root system is

$$\text{SO}(2l+1) : \mathfrak{R} = \{\pm\underline{e}_a \pm \underline{e}_b, a < b; \pm\underline{e}_a\} \tag{14.26}$$

The positive roots are immediately recognised:

$$\mathfrak{R}_+ = \{\underline{e}_a \pm \underline{e}_b, a < b; \underline{e}_a\} \tag{14.27}$$

What about the simple roots? Those SO$(2l)$ roots which were *not* simple will again be not simple, since SO$(2l) \subset$ SO$(2l+1)$. So the new simple roots must be some subset of the old ones, and the new positive roots. Thus we must search among

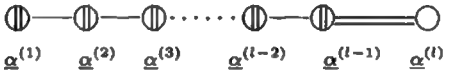
$$\underline{e}_1 - \underline{e}_2, \underline{e}_2 - \underline{e}_3, \ldots, \underline{e}_{l-1} - \underline{e}_l, \underline{e}_{l-1} + \underline{e}_l, \underline{e}_1, \underline{e}_2, \ldots, \underline{e}_l$$

and remove those that are expressible as positive integer combinations of others. In this way, we are able to eliminate $\underline{e}_{l-1} + \underline{e}_l, \underline{e}_1, \underline{e}_2, \ldots, \underline{e}_{l-1}$. The survivors are

the simple roots for SO(2l + 1):

$$\begin{aligned} \mathcal{S} &= \{\underline{e}_1 - \underline{e}_2, \underline{e}_2 - \underline{e}_3, \ldots, \underline{e}_{l-1} - \underline{e}_l, \underline{e}_l\}, \\ \underline{\alpha}^{(1)} &= \underline{e}_1 - \underline{e}_2, \underline{\alpha}^{(2)} = \underline{e}_2 - \underline{e}_3, \ldots, \underline{\alpha}^{(l-1)} = \underline{e}_{l-1} - \underline{e}_l, \\ \underline{\alpha}^{(l)} &= \underline{e}_l \end{aligned} \tag{14.28}$$

Notice that the geometry is different! All of $\underline{\alpha}^{(1)}, \underline{\alpha}^{(2)}, \ldots, \underline{\alpha}^{(l-1)}$ have length $\sqrt{2}$, and successive ones make an angle of 120° with one another (those two or more steps apart are perpendicular!). The last root $\underline{\alpha}^{(l)}$ has unit length, and makes an angle of 135° with $\underline{\alpha}^{(l-1)}$: all this agrees with the general angle-length ratio restrictions of the previous Chapter. Thus we obtain the Dynkin diagram for SO(2l + 1):
π-System for SO(2l + 1) ≡ B_l



(14.29)

The longer roots have been represented by shaded circles.

14.3 The USp(2l) Family — C_l of Cartan

The family of unitary symplectic groups, defined only in complex spaces with an even number of dimensions, is somewhat unfamiliar to most physicists. We therefore analyse its Lie algebra structure in a little detail. Curiously, the groups USp(2l) seem to lie "mid-way between" SO(2l) and SO(2l + 1), in some aspects resembling the former and in others the latter.

The defining representation D of USp(2l) consists of complex 2l-dimensional matrices U obeying two conditions-the *unitary* condition, and the *symplectic* condition. To set up the latter, we have to define a "symplectic metric". As with SO(2l), let indices $A, B, \ldots$ go over $1, 2, \ldots, 2l$; and use split index notation when needed. Then the symplectic metric is the 2l-dimensional real antisymmetric matrix

$$\eta = (\eta_{AB}) = \begin{pmatrix} 0 & 1 & & & & \\ -1 & 0 & & & & \\ & & 0 & 1 & & \\ & & -1 & 0 & & \\ & & & & & \\ & \bigcirc & & & \bigcirc & \ddots \end{pmatrix} = \begin{pmatrix} i\sigma_2 & & & \\ & i\sigma_2 & & \bigcirc \\ & & & \\ & \bigcirc & & \ddots \end{pmatrix}$$

$$\begin{aligned} &\eta_{AB} \equiv \eta_{ar,bs} = \delta_{ab}\epsilon_{rs}, \\ &(\epsilon_{rs}) = i\sigma_2 \end{aligned} \tag{14.30}$$

(It is because η has to be both antisymmetric and nonsingular that we must have an even number of dimensions). The important properties of η are:

$$\eta^2 = -1, \eta^T = \eta^\dagger = -\eta \tag{14.31}$$

Now the matrices $U \in D$ are defined by the two requirements

$$U^\dagger U = 1, \quad U^T \eta U = \eta \tag{14.32}$$

Let us examine the form of an U close to the identity:

$$U \simeq 1 - i\epsilon J, |\epsilon| \ll 1 : J^\dagger = J, J^T\eta + \eta J = 0 \tag{14.33}$$

The symplectic condition on J can also be expressed as

$$(\eta J)^T = \eta J \tag{14.34}$$

We must find in a convenient way the real independent parameters in a matrix J obeying these conditions; that will give us the order of the group, a basis for the Lie algebra etc. It is useful to recognise that there is a subgroup formed by the direct product of independent SU(2)'s acting on the pairs $12, 34, \ldots, 2l-1, 2l$:

$$\begin{array}{rccccc} & SU(2) \times SU(2) \times \ldots \times SU(2) \subset USp(2l) & & & & \\ A = 1,2 & & 3,4 & \ldots & 2l-1, 2l & \\ \text{i.e.,} \quad a = 1 & & 2 & \ldots & l & \end{array} \tag{14.35}$$

With the use of split indices, we break up J into 2×2 blocks as follows: On the main diagonal, we have $J^{(a)}$ at the a^{th} position, for $a = 1, 2, \ldots, l$. Then for $a \neq b$, we have $J^{(a,b)}$ at the intersection of the a^{th} pair of rows and b^{th} pair of columns. Then the picture is:

$$(J_{AB}) = \begin{array}{c} \\ \\ a \rightarrow \end{array} \begin{pmatrix} & & a & & b & \\ & & \downarrow & & \downarrow & \\ ( ) & & & & & \\ & \ddots & & & & \\ \ldots & & (J^{(a)}) & \ldots & (J^{(a,b)}) & \ldots \\ & \ddots & & & & \end{pmatrix} \tag{14.36}$$

One can immediately verify that the blocks in $J^\dagger, \eta J$ and $(\eta J)^T$ are related to those in J in the following ways:

$$\begin{aligned} (J^\dagger)^{(a)} &= J^{(a)\dagger}, & (J^\dagger)^{(a,b)} &= J^{(b,a)\dagger}; \\ (\eta J)^{(a)} &= i\sigma_2 J^{(a)}, & (\eta J)^{(a,b)} &= i\sigma_2 J^{(a,b)}; \\ ((\eta J)^T)^{(a)} &= (i\sigma_2 J^{(a)})^T, & ((\eta J)^T)^{(a,b)} &= (i\sigma_2 J^{(a,b)})^T \end{aligned} \tag{14.37}$$

Therefore, the conditions on J in Eq.(14.33) translate into these statements on the 2×2 blocks:

$$\begin{aligned} J^\dagger = J : \; & J^{(a)\dagger} = J^{(a)}, \quad J^{(a,b)\dagger} = J^{(b,a)}; \\ \eta J = \text{Symmetric} : \; & i\sigma_2 J^{(a)} = \text{Symmetric}, \quad i\sigma_2 J^{(a,b)} = (i\sigma_2 J^{(b,a)})^T \end{aligned} \tag{14.38}$$

For each of $J^{(a)}$ and $J^{(a,b)}$ we can introduce a complex linear combination of the 2×2 unit matrix and the Pauli matrices, and see how the coefficients get restricted. Then the result is that each $J^{(a)}$ is a real linear combination of the three Pauli matrices, symbolically,

$$J^{(a)} = x\sigma_1 + y\sigma_2 + z\sigma_3, \qquad xyz \text{ real}, \qquad a = 1, 2, \ldots, l \tag{14.39}$$

Thus these $J^{(a)}$'s are in fact the generators of the subgroup SU(2) $\times$ SU(2) $\times \ldots \times$ SU(2) of Eq.(14.35). Each $J^{(a,b)}$ for $a < b$ is a pure imaginary multiple of the unit matrix plus a real linear combination of the σ's:

$$\begin{aligned} a < b: \quad J^{(a,b)} &= x\sigma_1 + y\sigma_2 + z\sigma_3 + iw, \quad x, y, z, w \text{ real}; \\ J^{(b,a)} &= J^{(a,b)\dagger} \end{aligned} \tag{14.40}$$

The parameters in the $J^{(a,b)}$ for different $a < b$ pairs are of course independent. Counting up the $3l$ independent parameters in the $J^{(a)}$, and the $2l(l-1)$ parameters in the $J^{(a,b)}$, we see that there are in all $l(2l+1)$ independent parameters in J. Thus the order of USp(2l), as in the case of SO($2l+1$), is $l(2l+1)$.

The descriptions of $J^{(a)}$ and $J^{(a,b)}$ in terms of Pauli matrices guide us in choosing a basis for the Lie algebra. We take

$$X_j^{(a)} = \sigma_j \text{ at } a^{\text{th}} \text{ diagonal block} = \begin{pmatrix} \ddots & & \\ & \sigma_j & \\ & & \ddots \end{pmatrix}, \quad a = 1, 2, \ldots, l;$$

$$X_0^{(a,b)} = \pm i \cdot 1 \text{ at } ab \text{ and } ba \text{ blocks} = \begin{pmatrix} & \ddots & i1 \\ & & \\ -i1 & \ddots & \end{pmatrix}, \quad a < b = 1, 2, \ldots, l;$$

$$X_j^{(a,b)} = \sigma_j \text{ at } ab \text{ and } ba \text{ blocks} = \begin{pmatrix} \ddots & \sigma_j & \\ & & \\ \sigma_j & \ddots & \end{pmatrix}, \quad a < b = 1, 2, \ldots, l$$

(14.41)

The following subset of commutation relations follows quite easily:

$$\begin{aligned} [X_j^{(a)}, X_k^{(b)}] &= 2i\delta_{ab}\epsilon_{jkm}X_m^{(a)}; \\ b < c: [X_j^{(a)}, X_0^{(b,c)}] &= i(\delta_{ab} - \delta_{ac})X_j^{(b,c)}; \\ [X_j^{(a)}, X_k^{(b,c)}] &= i(\delta_{ab} - \delta_{ac})\delta_{jk}X_0^{(b,c)} \\ &\quad + i(\delta_{ab} + \delta_{ac})\epsilon_{jkm}X_m^{(b,c)}, \text{ no sums on } b, c \end{aligned} \tag{14.42}$$

These are all the relations involving at least one generator of SU(2) $\times$ SU(2) .... The remaining commutation relations to complete the Lie algebra structure are left as an exercise.

For a Cartan subalgebra, one might guess that the set $H_a = X_3^{(a)}$ would do. It turns out that this is so-these generators of the individual U(1)'s within each SU(2) do form a maximal abelian set of generators. Thus the rank of USp(2l) is seen to be l. Fortunately in the defining representation D the H_a are already diagonal, since H_a is simply σ_3 in the a^{th} diagonal block. So the weights in D are:

$$D \text{ of Usp}(2l)\text{: } \{\underline{\mu}\} = \{\pm\underline{e}_1, \pm\underline{e}_2, \ldots, \pm\underline{e}_l\} \tag{14.43}$$

There are $2l$ distinct weights, formally the same as in the SO(2l) case, and each is nondegenerate.

The total number of roots is order minus rank, so $2l^2$ as for SO(2l+1). The a^{th} factor in the product subgroup SU(2)× SU(2) × ... ×SU(2) already supplies us with the roots $\pm 2\underline{e}_a$: these correspond to the raising and lowering operators of that SU(2). There must then be $2l(l-1)$ additional roots.

These turn out to be $\pm\underline{e}_a \pm \underline{e}_b, a < b$, which incidentally were the complete set of roots for SO(2l). For the roots $\pm 2\underline{e}_a$, the E_α operators are easy:

$$[H_a, X_1^{(b)} \pm iX_2^{(b)}] = \pm 2\delta_{ab}(X_1^{(b)} \pm iX_2^{(b)}), \text{ no sum} \tag{14.44}$$

For the rest, we find from the partial list of commutation relations (14.42):

$$\begin{aligned} b < c : [H_a, X_0^{(b,c)} + i\epsilon X_3^{(b,c)}] &= \epsilon(\delta_{ab} - \delta_{ac})(X_0^{(b,c)} + i\epsilon X_3^{(b,c)}), \text{ no sum;} \\ [H_a, X_1^{(b,c)} + i\epsilon X_2^{(b,c)}] &= \epsilon(\delta_{ab} + \delta_{ac})(X_1^{(b,c)} + i\epsilon X_2^{(b,c)}), \text{ no sum;} \\ \epsilon &= \pm 1 \end{aligned} \tag{14.45}$$

The former therefore provide us with E_α's for roots of the form $\pm(\underline{e}_b - \underline{e}_c)$, the latter for roots of the form $\pm(\underline{e}_b + \underline{e}_c)$. The full list is then,

$$\text{USp}(2l)\text{: } \mathfrak{R} = \{\pm\underline{e}_a \pm \underline{e}_b, a < b; \pm 2\underline{e}_a\} \tag{14.46}$$

Positive roots are the subset of l^2 vectors

$$\mathfrak{R}_+ = \{\underline{e}_a \pm \underline{e}_b, \quad a < b; 2\underline{e}_a\} \tag{14.47}$$

In finding the set $\mathcal{S}$ of simple roots, we can drop those positive roots which in the case of SO(2l) were not simple. This is similar to the procedure we adopted for SO(2l + 1). Thus, now $\mathcal{S}$ is a subset of

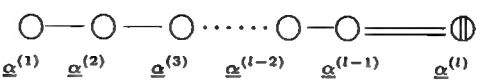
$$\underline{e}_1 - \underline{e}_2, \underline{e}_2 - \underline{e}_3, \ldots, \underline{e}_{l-1} - \underline{e}_l, \underline{e}_{l-1} + \underline{e}_l, 2\underline{e}_1, 2\underline{e}_2, \ldots, 2\underline{e}_l$$

The vectors that get eliminated at this stage are, as in the SO(2l + 1) case, $\underline{e}_{l-1} + \underline{e}_l, 2\underline{e}_1, 2\underline{e}_2, \ldots, 2\underline{e}_{l-1}$, so we are left with

$$\begin{aligned} \mathcal{S} &= \{\underline{e}_1 - \underline{e}_2, \underline{e}_2 - \underline{e}_3, \ldots, \underline{e}_{l-1} - \underline{e}_l, 2\underline{e}_l\}; \\ \underline{\alpha}^{(1)} &= \underline{e}_1 - \underline{e}_2, \underline{\alpha}^{(2)} = \underline{e}_2 - \underline{e}_3, \ldots, \underline{\alpha}^{(l-1)} = \underline{e}_{l-1} - \underline{e}_l, \\ \underline{\alpha}^{(l)} &= 2\underline{e}^{(l)} \end{aligned} \tag{14.48}$$

Notice the difference from the SO($2l+1$) case: Now the first $l-1$ simple roots each have length $\sqrt{2}$, the l^{th} has length 2, so the former are shorter. T he USp($2l$) Dynkin diagram follows:

π-system for Usp($2l$) $\equiv C_l$:



(14.49)

The difference as compared to SO($2l+1$) is that now only one circle is shaded: in Eq.(14.29), we have $(l-1)$ shaded circles.

It is illuminating to collect together the results of this section and the two previous ones, so that one can see at a glance the SO($2l$)– Usp($2l$)–SO($2l+1$) pattern (the common rank l is suppressed):

Group	*Order*	*dim. D*	$\{\underline{\mu}\}$ *in D*	$\mathfrak{R}$	$\mathcal{S}$
SO($2l$) $= D_l$	$l(2l-1)$	$2l$	$\pm\underline{e}_a$	$\pm\underline{e}_a \pm \underline{e}_b$	$\underline{e}_a - \underline{e}_{a+1}$; $\underline{e}_{l-1} + \underline{e}_l$
USp($2l$) $= C_l$	$l(2l+1)$	$2l$	$\pm\underline{e}_a$	$\pm\underline{e}_a \pm \underline{e}_b$, $\pm 2\underline{e}_a$	$\underline{e}_a - \underline{e}_{a+1}$; $2\underline{e}_l$
SO($2l+1$) $= B_l$	$l(2l+1)$	$2l+1$	$\pm\underline{e}_a, \underline{0}$	$\pm\underline{e}_a \pm \underline{e}_b$, $\pm\underline{e}_a$	$\underline{e}_a - \underline{e}_{a+1}$; $\underline{e}_l$

The ways in which USp($2l$) resembles SO($2l$), and those in which it is more like SO($2l+1$), are made evident. Of course intrinsically the symplectic groups are very different from the orthogonal ones, and for these the kind of intuitive Euclidean geometric ideas one is accustomed to are of no use.

14.4 The SU($l+1$) Family — A_l of Cartan

The algebraic work involved in analysing the defining representation D for SO($2l$), SO($2l+1$) and USp($2l$), and then deducing weights, roots etc., has been fairly clean. While the unitary groups SU($l+1$) are easier to define and picture, because of familiarity with quantum mechanics, they turn out to be in a numerical sense somewhat awkward when it comes to determination of roots, weights etc. Finally, of course, simplicity is regained in the Dynkin diagram!

The defining representation D of SU($l+1$) consists of all complex $(l+1)$-dimensional unitary unimodular matrices,

$$UU^\dagger = 1_{l+1}, \qquad \det U = +1. \tag{14.50}$$

As we know from experience in quantum mechanics, the infinitesimal generators are all hermitian traceless $(l+1)$-dimensional matrices (and it is the tracelessness that causes much of the ungainliness in the following expressions!). Since such a matrix involves $l(l+2)$ independent real parameters, that is the order of SU($l+1$).

There is a "tensor" way of describing the generators and Lie algebra of SU($l+1$), which we now describe but which we do not use later on. Let indices $j, k, m, \ldots$, run over $1, 2, \ldots, l+1$. Introduce the following set of matrices in $l+1$ dimensions:

$$\begin{aligned}(A^j{}_k)^m_n &= \delta^j_n \delta^m_k - \frac{1}{l+1}\delta^j_k \delta^m_n, \\ (A^j{}_k)^\dagger &= A^k{}_j, \\ A^j{}_j &= 0, \\ \mathrm{Tr} A^j{}_k &= 0 \end{aligned} \tag{14.51}$$

Any hermitian traceless $(l+1)$ dimensional matrix X, such that $U \simeq 1 - i\epsilon X$ is an infinitesimal element of SU($l+1$), can be uniquely written as a combination of these A's:

$$\begin{aligned} X &= x^k_j A^j{}_k, \\ x^{k*}_j &= x^j_k, \quad x^j_j = 0. \end{aligned} \tag{14.52}$$

Thus, the $A^j{}_k$ subject to the vanishing of $A^j{}_j$ can be used as a (nonreal) basis for the Lie algebra of SU($l+1$). Their commutation relations are quite simple:

$$[A^j{}_k, A^m{}_n] = \delta^j_n A^m_k - \delta^m_k A^j{}_n. \tag{14.53}$$

One can then define the Lie algebra of SU($l+1$) by these brackets, the hermiticity condition in Eq.(14.51), and the understanding that the $A^j{}_k$ are an *over complete* basis for the Lie algebra since the sum $A^j{}_j$ must vanish.

While one can certainly proceed with such a scheme, we shall instead start afresh and identify the Cartan subalgebra, weights and roots in a different fashion. In the defining representation D, the subgroup of diagonal matrices can be taken as the source of the H_a: thus we need a complete independent set of real diagonal $(l+1)$ dimensional traceless matrices. That these will be maximal is obvious. Any X commuting with all diagonal traceless matrices is itself necessarily diagonal. We therefore define the elements of the Cartan subalgebra in

D as follows (the vanishing off diagonal elements are not indicated):

$$H_1 = \frac{1}{\sqrt{1\cdot 2}} \begin{pmatrix} 1 & & & & \\ & -1 & & & \\ & & 0 & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix},$$

$$H_2 = \frac{1}{\sqrt{2\cdot 3}} \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & -2 & & \\ & & & 0 & \\ & & & & \ddots \end{pmatrix}, \ldots,$$

$$H_a = \frac{1}{\sqrt{a(a+1)}} \begin{pmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & -a & & \\ & & & & & 0 & \\ & & & & & & \ddots \end{pmatrix}, \ldots,$$

$$H_l = \frac{1}{\sqrt{l(l+1)}} \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & -l \end{pmatrix} \tag{14.54}$$

The rank of SU($l+1$) is verified to be l. These matrices are basically the $A^1{}_1, A^2{}_2, \ldots, A^{l+1}_{l+1}$ of the tensor description suitably rearranged and recombined; the numerical factors have been chosen so that *in the defining representation* D, we get

$$\text{Tr } H_a H_b = \delta_{ab}. \tag{14.55}$$

For dealing with the weights $\{\mu\}$ in D and later in finding the E_α combinations, it helps to use quantum mechanical ket vector notation for the $l+1$ basic states on which the SU($l+1$) matrices $U \in D$ are made to act. Thus we shall have the kets $|j\rangle$ for $j = 1, 2, \ldots, l+1$. Since the H_a in Eq.(14.54) are already all diagonal, the weights occurring and the corresponding eigenvectors can all

be directly read off. We give the expressions in some detail:

$$H_a|j\rangle = \mu_a^{(j)}|j\rangle, \text{ no sum, } a = 1,2,\ldots,l; j = 1,2,\ldots,l+1;$$

$$\underline{\mu}^{(1)} = \left(\frac{1}{\sqrt{1.2}}, \frac{1}{\sqrt{2.3}}, \frac{1}{\sqrt{3.4}}, \ldots, \frac{1}{\sqrt{(l-1)l}}, \frac{1}{\sqrt{l(l+1)}}\right);$$

$$\underline{\mu}^{(2)} = \left(\frac{-1}{\sqrt{1.2}}, \frac{1}{\sqrt{2.3}}, \frac{1}{\sqrt{3.4}}, \ldots, \frac{1}{\sqrt{(l-1)l}}, \frac{1}{\sqrt{l(l+1)}}\right);$$

$$\underline{\mu}^{(3)} = \left(0, \frac{-2}{\sqrt{2.3}}, \frac{1}{\sqrt{3.4}}, \ldots, \frac{1}{\sqrt{(l-1)l}}, \frac{1}{\sqrt{l(l+1)}}\right);$$

$$\underline{\mu}^{(4)} = \left(0, 0, \frac{-3}{\sqrt{3.4}}, \ldots, \frac{1}{\sqrt{(l-1)l}}, \frac{1}{\sqrt{l(l+1)}}\right);$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\underline{\mu}^{(l)} = \left(0, 0, 0, \ldots, \frac{-(l-1)}{\sqrt{(l-1)l}}, \frac{1}{\sqrt{l(l+1)}}\right);$$

$$\underline{\mu}^{(l+1)} = \left(0, 0, 0, \ldots, 0, \frac{-l}{\sqrt{l(l+1)}}\right) \tag{14.56}$$

The weights $\underline{\mu}^{(j)}, j = 1,2,\ldots,l+1$ of the defining representation D are all distinct and nondegenerate.

To find the roots and E_α combinations, we take the remaining generators of SU$(l+1)$ to be:

$$\begin{aligned}(A_{jk})_{mn} &= \delta_{jm}\delta_{kn} + \delta_{jn}\delta_{km}, \quad j \neq k: \\ (B_{jk})_{mn} &= -i(\delta_{jm}\delta_{kn} - \delta_{jn}\delta_{km}), \quad j \neq k\end{aligned} \tag{14.57}$$

These are of course not independent, though they are hermitian:

$$\begin{aligned}A_{jk}^\dagger &= A_{jk} = A_{kj} \\ B_{jk}^\dagger &= B_{jk} = -B_{kj}\end{aligned} \tag{14.58}$$

Thus, there are $\frac{1}{2}l(l+1)A$'s and an equal number of B's which are independent. The E_α's happen to be the complex combinations,

$$\begin{aligned}E_{jk} &= \frac{1}{2}(A_{jk} + iB_{jk}), \quad j \neq k; \\ (E_{jk})_{mn} &= \delta_{jm}\delta_{kn}\end{aligned} \tag{14.59}$$

Thus, (in the representation D), E_{jk} has only one non-zero matrix element, at the j^{th} row k^{th} column. Since the actions of both H_a and E_{jk} on the kets $|j\rangle$ are extremely simple, the roots are easily obtained:

$$\begin{aligned}&H_a|j\rangle = \underline{\mu}_a^{(j)}|j\rangle, \; E_{jk}|m\rangle = \delta_{km}|j\rangle \;\Rightarrow \\ &[H_a, E_{jk}] = (\underline{\mu}_a^{(j)} - \underline{\mu}_a^{(k)})E_{jk}, j \neq k, \qquad \text{no sum.}\end{aligned} \tag{14.60}$$

Therefore, the roots are *all* differences of weights of the defining representation!

$$\mathrm{SU}(l+1):\mathfrak{R}=\{\underline{\mu}^{(j)}-\underline{\mu}^{(k)},j\neq k;j,k=1,2,\ldots,l+1\} \tag{14.61}$$

From here to find the positive roots, and then the simple ones, one has to examine in detail the weight vector components listed in Eq.(14.56). The positive roots are found to be the following:

$$\begin{aligned}\mathfrak{R}_+ = &\ \underline{\mu}^{(1)}-\underline{\mu}^{(j)},\ j=2,3,\ldots,l,l+1;\\ &\ \underline{\mu}^{(3)}-\underline{\mu}^{(j)},\ j=2;\\ &\ \underline{\mu}^{(4)}-\underline{\mu}^{(j)},\ j=2,3;\\ &\ \ldots\ldots\ldots\ldots\ldots\\ &\ \underline{\mu}^{(l)}-\underline{\mu}^{(j)},\ j=2,3,\ldots,l-1;\\ &\ \underline{\mu}^{(l+1)}-\underline{\mu}^{(j)},\ j=2,3,\ldots,l.\end{aligned} \tag{14.62}$$

The number of vectors here is $\frac{1}{2}l(l+1)$, exactly half the number in $\mathfrak{R}$. A more concise way of describing $\mathfrak{R}_+$ is this:

$$\mathfrak{R}_+=\{\underline{\mu}^{(1)}-\underline{\mu}^{(j)},j\geq 2;\underline{\mu}^{(j)}-\underline{\mu}^{(k)},j>k\geq 2\} \tag{14.63}$$

Out of these positive roots to pick out the simple ones is again a bit tedious, but by looking at small values of l, like $l=2,3,4,\ldots$, one quickly sees the pattern. The general result is:

$$\mathcal{S}=\underline{\mu}^{(l+1)}-\underline{\mu}^{(l)},\underline{\mu}^{(l)}-\underline{\mu}^{(l-1)},\ldots,\underline{\mu}^{(3)}-\underline{\mu}^{(2)},\underline{\mu}^{(1)}-\underline{\mu}^{(l+1)} \tag{14.64}$$

We prefer to rearrange them in the following sequence (remember that each root has l components):

$$\begin{aligned}
\underline{\alpha}^{(1)} &= \underline{\mu}^{(3)}-\underline{\mu}^{(2)}=\left(\frac{1}{\sqrt{1.2}},\frac{-3}{\sqrt{2.3}},0,\ldots,0\right),\\
\underline{\alpha}^{(2)} &= \underline{\mu}^{(4)}-\underline{\mu}^{(3)}=\left(0,\frac{2}{\sqrt{2.3}},\frac{-4}{\sqrt{3.4}},0,\ldots,0\right),\\
\underline{\alpha}^{(3)} &= \underline{\mu}^{(5)}-\underline{\mu}^{(4)}=\left(0,0,\frac{3}{\sqrt{3.4}},\frac{-5}{\sqrt{4.5}},0,\ldots,0\right),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
\underline{\alpha}^{(l-2)} &= \underline{\mu}^{(l)}-\underline{\mu}^{(l-1)}=\left(0,\ldots,0,\frac{l-2}{\sqrt{(l-2)(l-1)}},\frac{-l}{\sqrt{(l-1)l}},0\right),\\
\underline{\alpha}^{(l-1)} &= \underline{\mu}^{(l+1)}-\underline{\mu}^{(l)}=\left(0,\ldots,0,\frac{l-1}{\sqrt{(l-1)l}},\frac{-(l+1)}{\sqrt{l(l+1)}}\right),\\
\underline{\alpha}^{(l)} &= \underline{\mu}^{(1)}-\underline{\mu}^{(l+1)}=\left(\frac{1}{\sqrt{1.2}},\frac{1}{\sqrt{2.3}},\frac{1}{\sqrt{3.4}},\ldots,\frac{1}{\sqrt{(l-1)l}},\frac{l+1}{\sqrt{l(l+1)}}\right);\\
\mathcal{S} &= \{\underline{\alpha}^{(a)},a=1,2,\ldots,l\}
\end{aligned} \tag{14.65}$$

These algebraic complications in the case of SU($l+1$) seem pretty well unavoidable; the observance of the tracefree condition is much like working with independent variables in the centre of mass frame of a many-particle system!

To draw the Dynkin diagram, we need to compute the lengths and angles among simple roots. Fortunately at this stage things are simple:

$$|\underline{\alpha}^{(a)}| = \sqrt{2}, a = 1, 2, \ldots, l;$$
$$\underline{\alpha}^{(a)} \cdot \underline{\alpha}^{(a+1)} = -1, a = 1, 2, \ldots, l-1, \text{ rest zero} \tag{14.66}$$

Therefore, the Dynkin diagram is much simpler than in the three earlier cases:

π-system for SU($l+1$) $\equiv A_l$:

$$\begin{array}{ccccccccc} \circ & - & \circ & - & \circ & \cdots & - & \circ & - & \circ \\ \underline{\alpha}^{(1)} & & \underline{\alpha}^{(2)} & & \underline{\alpha}^{(3)} & \cdots & & \underline{\alpha}^{(l-1)} & & \underline{\alpha}^{(l)} \end{array} \tag{14.67}$$

14.5 Coincidences for low Dimensions and Connectedness

We have described in some detail the defining representations of the four classical families of compact simple Lie groups $A_l = \mathrm{SU}(l+1), B_l = \mathrm{SO}(2l+1), C_l = \mathrm{USp}(2l), D_l = \mathrm{SO}(2l)$. These representations were exploited to find the Cartan subalgebras, roots, positive roots, simple roots and then to draw the π-systems. For small values of l, there are "chance" coincidences among the Lie algebras of these four families, after which they branch out in "independent directions". As good a way as any to spot these coincidences is to look at the respective Dynkin diagrams, thus putting them to use! In this way one finds the local isomorphisms:

$$\begin{aligned} A_1 \sim B_1 \sim C_1 &: \ \mathrm{SU}(2) \sim \mathrm{SO}(3) \sim \mathrm{USp}(2) \\ B_2 \sim C_2 &: \ \mathrm{SO}(5) \sim \mathrm{USp}(4) \\ A_3 \sim D_3 &: \ \mathrm{SU}(4) \sim \mathrm{SO}(6) \end{aligned} \tag{14.68}$$

Beyond these, as the diagrams show, there are no more coincidences.

As given by their defining representations the groups A_l and C_l turn out to be simply connected, and so they are their own universal covering groups. On the other hand, both B_l and D_l are doubly connected, so their universal covering groups give in each case a two-fold covering of the group specified by the defining representation.

Exercises for Chapter 14

1. From the commutation relations (14.6) in the case of SO(4) corresponding to $l = 2$, show that the Lie algebra splits into two mutually commuting SO(3) Lie algebras.

2. For the defining representation of the group SU(3), the analogues of the three Pauli matrices $\underline{\sigma}$ familiar from SU(2) are the eight GellMann matrices λ_r defined as follows:

$$\lambda_1 = \begin{pmatrix} \sigma_1 & & 0 \\ & & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \lambda_2 = \begin{pmatrix} \sigma_2 & & 0 \\ & & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \lambda_3 = \begin{pmatrix} \sigma_3 & & 0 \\ & & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \ \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \ \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \ \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix},$$

Verify the commutation and anticommutation relations

$$[\lambda_r, \lambda_s] = 2i f_{rst} \lambda_t, \{\lambda_r, \lambda_s\} = \frac{4}{3}\delta_{rs} + 2d_{rst}\lambda_t$$

where the completely antisymmetric f_{rst} and completely symmetric d_{rst} have these independent nonzero values:

$$f_{123} = 1, f_{147} = -f_{156} = f_{246} = f_{257} = f_{345} = -f_{367} = \frac{1}{2},$$

$$f_{458} = f_{678} = \frac{\sqrt{3}}{2};$$

$$d_{118} = d_{228} = d_{338} = -d_{888} = \frac{1}{\sqrt{3}},$$

$$d_{146} = d_{157} = -d_{247} = d_{256} = d_{344} = d_{355} = -d_{366} = -d_{377} = \frac{1}{2},$$

$$d_{448} = d_{558} = d_{668} = d_{778} = -\frac{1}{2\sqrt{3}}.$$

3. Verify explicitly at the Lie algebra level the low-dimensional coincidences listed in Eq.(14.68).

Chapter 15

Complete Classification of All CSLA Simple Root Systems

In Chapters 12 and 13, the general geometrical properties of roots, positive and simple, were derived. These properties are comprehensive enough to allow us to determine exhaustively all possible simple root systems. The detailed look at A_l, B_l, C_l and D_l in the previous Chapter has given us examples of the general theory in action. Now we take the opposite point of view and use the geometrical restrictions to completely determine all allowed simple root systems. The question naturally is: What are the possibilities apart from those encountered in the last Chapter? As promised, we will be showing that there are only five other CSLA's, the so-called *exceptional* algebras.

Let us right away list the three basic conditions that the system $\mathcal{S}$ of simple roots of a CSLA of rank l must obey:

(A) $\mathcal{S}$ consists of l independent vectors in l-dimensional real Euclidean space.

(B) The possible angles and length ratios among these vectors are 90° (free ratio), 120° (ratio 1), 135° (ratio $\sqrt{2}$) and 150° (ratio $\sqrt{3}$).

(C) The system $\mathcal{S}$ must be indecomposable, i.e., it must not split into two disjoint subsets with every vector in one being perpendicular to every vector in the other. This connectedness of a Dynkin diagram was proved in Chapter 13.

Throughout this Chapter, we shall frequently refer to these three conditions (A), (B), (C). We have called a system $\mathcal{S}$ satisfying (A), (B) and (C) a π-system. An allowed π-system or Dynkin diagram will often be given the symbol Γ, and be referred to as a graph; a subgraph will often be denoted by Γ', etc.

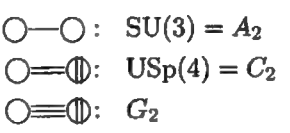
15.1 Series of Lemmas

Now we present a carefully ordered sequence of lemmas which, little by little, narrow down the possible π-systems Γ.

Lemma I: A *connected* subgraph Γ' of an allowed graph Γ is also allowed. For, properties (A) and (B) follow for Γ' from their validity for Γ; and (C) is ensured for Γ' by insisting that it be connected. Of course, l' for Γ' is less than l for Γ.

Lemma II. For $l = 1$, the only possible Γ is O, corresponding to the group SU(2)= A_1; and this is certainly allowed.

For $l = 2$, the possible Γ's are



(15.1)

Possibilities A_2 and C_2 are known to us; the third, the Lie algebra of the group G_2, is also allowed. It is a group of order 14 and rank 2, the smallest of the five exceptional groups. We shall exhibit its roots later in this Chapter.

Lemma III: For $l = 3$, the only possible Γ's are:

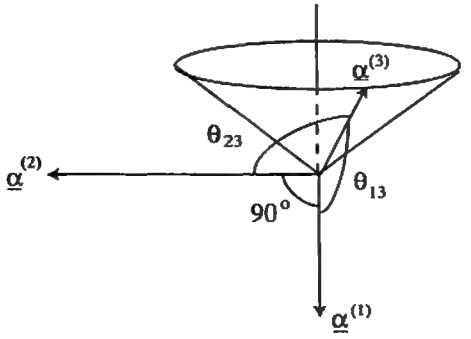
(15.2)

corresponding to SU(4), SO(7) and USp(6). The proof is as follows.

The three vectors $\underline{\alpha}^{(1)}, \underline{\alpha}^{(2)}, \underline{\alpha}^{(3)}$ must be linearly independent. If the enclosed angles are $\theta_{12}, \theta_{23}, \theta_{31}$, we can agree to order them as $\theta_{12} \leq \theta_{23} \leq \theta_{31}$. Each θ is either 90° or 120° or 135° or 150°. Now the linear independence condition gives this restriction:

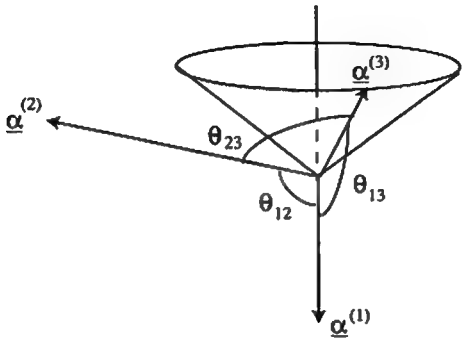
$$\theta_{12} + \theta_{23} + \theta_{13} < 360° \tag{15.3}$$

This is best seen pictorially. Suppose $\theta_{12} = 90°$; then $\theta_{23}, \theta_{13} > 90°$ (Recall that property (C) forbids *two* of the θ's being 90°!). Whatever θ_{13} may be,



$\underline{\alpha}^{(3)}$ must lie on a cone with $-\underline{\alpha}^{(1)}$ as inner axis and semi inner angle $180° - \theta_{13}$. By moving the possible location of $\underline{\alpha}^{(3)}$ along the cone we see that the maximum

value of θ_{23} occurs *exactly* when the three vectors are coplanar, and that value is $360° - 90° - \theta_{13}$. Since they cannot be coplanar, the inequality (15.3) is proved in the case $\theta_{12} = 90°$. If $\theta_{12} > 90°$, then $\theta_{13} \geq \theta_{12}$, only a slightly different picture needs to be drawn, but the conclusion is the same:



Therefore, for $l = 3$, property (A) has given us the inequality (15.3), while property (C) implies that at most *one* of the θ's can be 90°. The possible values are then

$$(\theta_{12}, \theta_{23}, \theta_{13}) = (90°, 120°, 120°), (90°, 120°, 135°) \qquad (15.4)$$

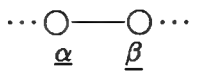
and these just correspond to the graphs (15.2). (Check that the configuration $(120°, 120°, 120°)$ with graph  fails to obey (A))

Lemma IV. On combining Lemma I with Lemma III, we can say that any connected set of three vectors in an allowed Γ for $l \geq 4$ must be of one of the two types shown in (15.2). That is, if we take any three connected vectors in Γ, drop all their connections to *other* vectors in Γ and "rub out" those other vectors, what remains must be one of the two diagrams (15.2).

Lemma V. We notice that in the two possible $l = 3$ cases (15.2), a triple line connection nowhere appears. This is due, of course, to the inequality (15.3). One can now combine Lemma IV and property (C) to state: in any allowed Γ for $l \geq 4$, the triple line connection given below

can *never appear.* Since it is absent for $l = 3$ as well, the only allowed Γ containing a triple line is at $l = 2$ for G_2, seen in Lemma II. Hereafter, then, *we can ignore such connections (angle = 150°) completely.*

Lemma VI. If in an allowed graph Γ, we have two circles connected by a *single* line, whatever else there may be, we can shrink this line, replace the two roots by a simple root of the same length, and get a graph Γ' which is also allowed. To prove this, let Γ look like,



By lemma IV, there is *no* vector $\underline{\gamma}$ which is connected to both $\underline{\alpha}$ and $\underline{\beta}$. So for any $\underline{\gamma} \in \Gamma$ distinct from both $\underline{\alpha}$ and $\underline{\beta}$, we have three choices:

(i) $\underline{\gamma} \cdot \underline{\alpha} = \underline{\gamma} \cdot \underline{\beta} = 0$: $\underline{\gamma}$ not connected to either $\underline{\alpha}$ or $\underline{\beta}$,
(ii) $\underline{\gamma} \cdot \underline{\alpha} \neq 0, \underline{\gamma} \cdot \underline{\beta} = 0$: $\underline{\gamma}$ connected to $\underline{\alpha}$, not to $\underline{\beta}$,
(iii) $\underline{\gamma} \cdot \underline{\alpha} = 0, \underline{\gamma} \cdot \underline{\beta} \neq 0$: $\underline{\gamma}$ connected to $\underline{\beta}$, not to $\underline{\alpha}$.

Now since the angle between α and β is 120°,

$$|\underline{\alpha}|^2 = |\underline{\beta}|^2 = |\underline{\alpha} + \underline{\beta}|^2$$

If we construct Γ' from Γ by coalescing the two circles and replacing the two vectors $\underline{\alpha}$ and $\underline{\beta}$ by $\underline{\alpha} + \underline{\beta}$, then in each of the three situations above we have:

(i) $\underline{\gamma} \cdot (\underline{\alpha} + \underline{\beta}) = 0$: $\underline{\gamma}$ remains unconnected to new $\underline{\alpha} + \underline{\beta}$;
(ii) $\underline{\gamma} \cdot (\underline{\alpha} + \underline{\beta}) = \underline{\gamma} \cdot \underline{\alpha}$: $\underline{\gamma}$ is now connected to $\underline{\alpha} + \underline{\beta}$ as it previously was to $\underline{\alpha}$, with no change in angles and lengths;
(iii) Interchange $\underline{\alpha}$ and $\underline{\beta}$ in (ii) just above.

Let us then check the properties (A), (B), (C) for Γ'. Property (A) holds because it was true for Γ; the rank however is reduced by one. Property (B) has just been shown, and (C) holds by the way we constructed Γ'. So, Γ' is an allowed graph with rank one less than for Γ.

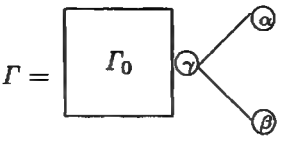
As we will see this is indeed a very powerful result. It allows us to come "cascading down" from an arbitrarily large graph to a graph of low rank, then apply our earlier results on the allowed low order graphs.

Lemma VII. The number of double lines in an allowed Γ is *at most one.* For $l = 3$, we have this in Lemma III. For $l \geq 4$: if an allowed Γ has two or more double lines, we can repeatedly use Lemma VI to keep on shrinking all single lines, at each stage obtaining an allowed Γ', until we end up with an allowed Γ'' with $l \geq 3$ and made up of *double lines only.* But this violates Lemma IV! So, either an allowed Γ has no double lines, or just one double line.

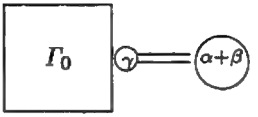
Lemma VIII. An allowed graph Γ cannot have any closed loops. For, if it does, then by Lemma IV, the number of vectors in the loop must be more than three. Then by Lemma VII, all lines in this loop *except possibly one* are single lines. Now use Lemma VI to keep shrinking these single lines one by one, at each stage getting an allowed graph. You will end up with a loop with three vectors, which Lemma IV disallows! Note that after each use of Lemma VI, the result would have had to be an allowed graph with a loop.

At this stage, then, we can say: an allowed Γ has only single lines and some branch points; or one double line, remaining single lines, and some branch points; no loops in either case. The next two Lemmas tackle the branch points.

Lemma IX. If an allowed graph Γ has the appearance



then, (i)



is allowed;

(ii) Γ_0 has no double lines inside itself. The proof is quite cute. By Lemma I, the subgraph Γ_0 (including γ) is an allowed one. Now for the three vectors $\underline{\alpha}, \underline{\beta}, \underline{\gamma}$ in Γ, we have

$$|\underline{\alpha}| = |\underline{\beta}| = |\underline{\gamma}|$$
$$\underline{\alpha} \cdot \underline{\beta} = 0$$
$$\underline{\alpha} \cdot \underline{\gamma} = \underline{\beta} \cdot \underline{\gamma} = -\frac{1}{2}|\underline{\alpha}|^2$$

Therefore,

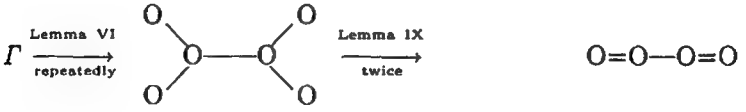
$$|\underline{\alpha} + \underline{\beta}| = \sqrt{2}|\underline{\gamma}|, \quad \theta_{\gamma,\alpha+\beta} = 135°.$$

For the graph Γ' built up above, (A) is valid because it is so for Γ; (B) holds by the above calculation; (C) holds by the construction. We see that Γ' is an allowed graph. Then by Lemma VII, Γ_0 cannot have had any double lines!

Lemma X. An allowed Γ can have at most one branch point. If it has one, then it cannot have any double lines. (The number of branches at the branch point will be controlled by Lemma XI!)

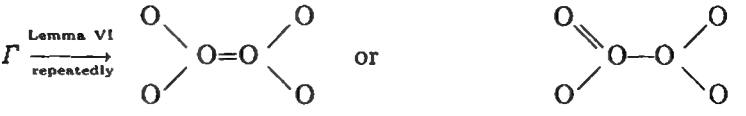
The proof is quickest by pictures. We have three cases to consider:

(a) If Γ has two or more branch points and no double lines, then



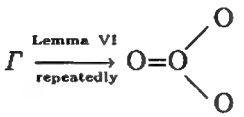
Violates Lemma VII

(b) If Γ has two or more branch points and one double line, then



both violate Lemma IX

(c) If Γ has one branch point and one double line, then

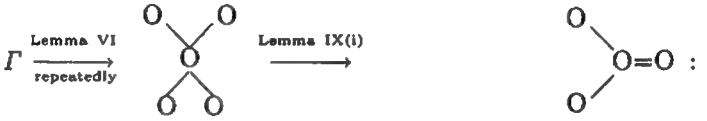


: Violates Lemma IX

In all the above, no assumption was made about the number of branches in Γ at a branch point.

So we have reached the point where we can say: in an allowed Γ, the number of double lines and the number of branch points cannot each exceed one, and the sum of these two numbers also cannot exceed one! We have next the last Lemma.

Lemma XI. If an allowed Γ has a branch point (hence no double lines) the number of branches there must be exactly three. For, if there were four or more branches, then



Violates Lemma IX (ii)

These Lemmas have drastically reduced the graphs Γ to be examined. Let us see next what the survivors are like!

15.2 The allowed Graphs Γ

We can classify the different possibilities that remain according to the number of branch points and double lines. There are then three situations.

(x) *No branch point, no double line*: This leads to the graph,

$$\Gamma = \underset{1}{O}\text{—}\underset{2}{O}\text{—}\underset{3}{O}\cdots O\text{—}\underset{l-1}{O}\text{—}\underset{l}{O}$$

which is certainly a possible graph, in fact the one for $A_l = \mathrm{SU}(l+1)$, Eq.(14.67).

(y) *No branch point, one double line*: One has the general possibility,

$$\Gamma = \underset{1}{\overset{\sqrt{2}}{O}}\text{—}\underset{2}{\overset{\sqrt{2}}{O}}\text{—}\underset{3\cdots l_1}{\overset{\sqrt{2}}{O}}\cdots O\overset{\sqrt{2}}{=}\underset{l_1+1}{\overset{1}{O}}\text{———}\underset{l_1+2}{\overset{1}{O}}\cdots O\text{———}\underset{l_1+l_2=l}{\overset{1}{O}}$$

with $l_1(l_2)$ vectors before (after) the double line. We assume that the former have lengths $\sqrt{2}$, the latter lengths unity, as indicated in the drawing. So to begin with we have

$$l_1 \geq 1, \quad l_2 \geq 1, \quad l = l_1 + l_2.$$

Now the condition (A) of linear independence of simple roots comes into the act and puts limits on l_1 and l_2, It shows that $l_1 \geq 2, l_2 \geq 3$ and $l_1 \geq 3, l_2 \geq 2$ are both disallowed (these are overlapping regimes!). If we had $l_1 \geq 2, l_2 \geq 3$ to begin with, then such a Γ reduces by repeated use of Lemma VI to $l_1 = 2, l_2 = 3$:

$$\begin{array}{rc} & \sqrt{2}\ \ \sqrt{2}\quad\ 1\quad\ 1\quad\ 1 \\ \Gamma' = & \mathrm{O}\text{—}\mathrm{O}\text{═}\mathrm{O}\text{—}\mathrm{O}\text{—}\mathrm{O} \\ & \underline{\alpha}^{(1)}\ \ \underline{\alpha}^{(2)}\ \ \underline{\alpha}^{(3)}\ \ \underline{\alpha}^{(4)}\ \ \underline{\alpha}^{(5)} \end{array}$$

But condition (A) is violated because as one can easily calculate,

$$|\underline{\alpha}^{(1)} + 2\underline{\alpha}^{(2)} + 3\underline{\alpha}^{(3)} + 2\underline{\alpha}^{(4)} + \underline{\alpha}^{(5)}|^2 = 0.$$

Therefore, along with Γ', even Γ is disallowed. With $l_1 \geq 3, l_2 \geq 2$, we would have reduced Γ to $l_1 = 3, l_2 = 2$:

$$\begin{array}{rc} & \sqrt{2}\ \ \sqrt{2}\quad\ \sqrt{2}\quad\ 1\quad\ 1 \\ \Gamma'' = & \mathrm{O}\text{—}\mathrm{O}\text{—}\mathrm{O}\text{═}\mathrm{O}\text{—}\mathrm{O} \\ & \underline{\alpha}^{(1)}\ \ \underline{\alpha}^{(2)}\ \ \underline{\alpha}^{(3)}\ \ \underline{\alpha}^{(4)}\ \ \underline{\alpha}^{(5)} \end{array}$$

and now

$$|\underline{\alpha}^{(1)} + 2\underline{\alpha}^{(2)} + 3\underline{\alpha}^{(3)} + 4\underline{\alpha}^{(4)} + 2\underline{\alpha}^{(5)}|^2 = 0.$$

What then remains in the no branch point, one double line case? We have: $l_1 = 1, l_2 \geq 1$:

$$\begin{array}{l} \sqrt{2}\quad 1\quad 1\quad 1\quad 1 \\ \mathrm{O}\text{═}\mathrm{O}\text{—}\mathrm{O}\cdots\mathrm{O}\text{—}\mathrm{O}: \ l = l_2 + 1 \geq 2 : \mathrm{USp}(2l) = C_l, \\ \quad\longleftarrow\!\!\!- l_2 -\!\!\!\longrightarrow \end{array} \qquad \text{Eq.(14.49)}$$

$l_1 = 2, l_2 = 1$:

$$\begin{array}{l} \sqrt{2}\ \ \sqrt{2}\ \ 1 \\ \mathrm{O}\text{—}\mathrm{O}\text{═}\mathrm{O}: \ l = 3 : \mathrm{SO}(7) = B_3, \text{Eq.(14.29)} \end{array}$$

$l_1 = 2, l_2 = 2$:

$$\begin{array}{l} \sqrt{2}\ \ \sqrt{2}\quad 1\quad 1 \\ \mathrm{O}\text{—}\mathrm{O}\text{═}\mathrm{O}\text{—}\mathrm{O}: \ l = 4 : \text{Exceptional group } F_4, \text{ see below} \end{array}$$

$l_1 \geq 3, l_2 = 1$:

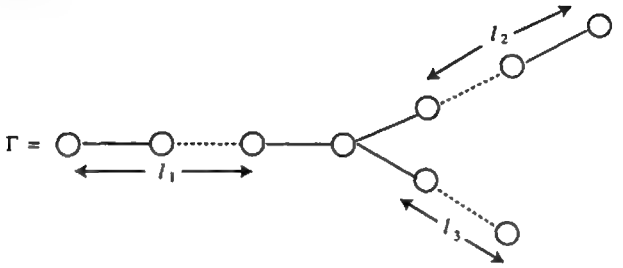
$$\begin{array}{l} \sqrt{2}\ \ \sqrt{2}\ \ \sqrt{2}\ \ \sqrt{2}\ \ 1 \\ \mathrm{O}\text{—}\mathrm{O}\cdots\mathrm{O}\text{—}\mathrm{O}\text{═}\mathrm{O}: \ l = l_1 + 1 \geq 4 : \mathrm{SO}(2l+1) = B_l, \\ \longleftarrow\!\!\!- l_1 -\!\!\!\longrightarrow \end{array} \qquad \text{Eq.(14.29)}$$

Therefore the no branch point-one double line case has given us USp($2l$) for $l \geq 2$, SO($2l+1$) for $l \geq 3$, and F_4.

(z) *One branch point, no double line*: Leaving out the circle at the vertex let the three arms have l_1, l_2, l_3 circles respectively, with

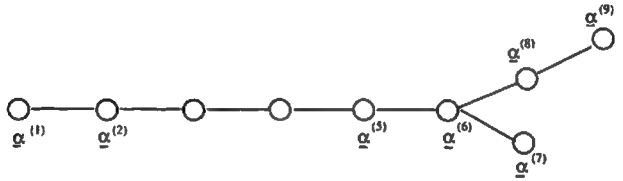
$$l_1 \geq l_2 \geq l_3 \geq 1, \quad l = l_1 + l_2 + l_3 + 1 \geq 4$$

So to begin we have the graph



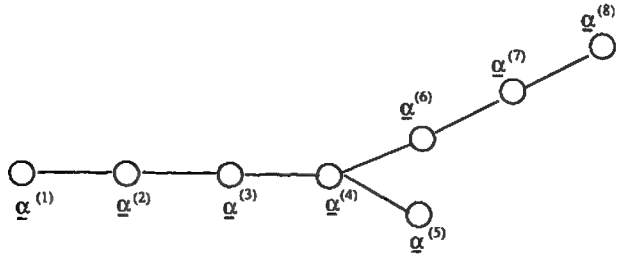
Our experience in situation (y) previously discussed suggests that if the arms are "excessively long", property (A) of linear independence of simple roots may fail. Indeed this is so. It happens that:

$$\Gamma(l_1 \geq 5, l_2 \geq 2, l_3) \xrightarrow[\text{VI}]{\text{Lemma}} \Gamma'(5,2,1) =$$



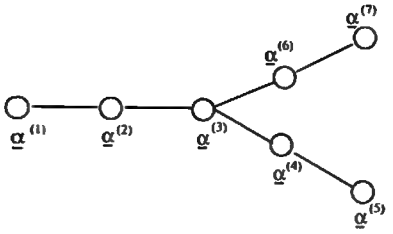
$$|\underline{\alpha}^{(1)} + 2\underline{\alpha}^{(2)} + 3\underline{\alpha}^{(3)} + 4\underline{\alpha}^{(4)} + 5\underline{\alpha}^{(5)} + 6\underline{\alpha}^{(6)} + 3\underline{\alpha}^{(7)} + 4\underline{\alpha}^{(8)} + 2\underline{\alpha}^{(9)}|^2 = 0.$$

$$\Gamma(l_1 \geq 3, l_2 \geq 3, l_3) \xrightarrow[\text{VI}]{\text{Lemma}} \Gamma'(3,3,1) =$$



$$|\underline{\alpha}^{(1)} + 2\underline{\alpha}^{(2)} + 3\underline{\alpha}^{(3)} + 4\underline{\alpha}^{(4)} + 2\underline{\alpha}^{(5)} + 3\underline{\alpha}^{(6)} + 2\underline{\alpha}^{(7)} + \underline{\alpha}^{(8)}|^2 = 0.$$

$$\Gamma(l_1 \geq 2, l_2 \geq 2, l_3 \geq 2) \xrightarrow[\text{VI}]{\text{Lemma}} \Gamma'(2,2,2) =$$



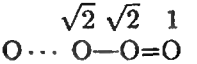
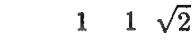
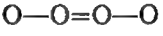
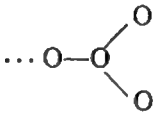
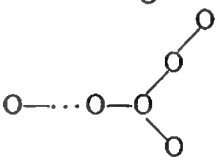
$$|\underline{\alpha}^{(1)} + 2\underline{\alpha}^{(2)} + 3\underline{\alpha}^{(3)} + 2\underline{\alpha}^{(4)} + \underline{\alpha}^{(5)} + 2\underline{\alpha}^{(6)} + \underline{\alpha}^{(7)}|^2 = 0.$$

After this large scale elimination of possibilities, the residue is the following:

l_1	l_2	l_3	*π-system for*
≥ 5	1	1	$l = l_1 + 3$: $\mathrm{SO}(2l) = D_l$ for $l \geq 8$
4	2	1	$l = 8$: Exceptional group E_8 (see below)
4	1	1	$l = 7$: $\mathrm{SO}(14) = D_7$
3	2	1	$l = 7$: Exceptional group E_7 (see below)
3	1	1	$l = 6$: $\mathrm{SO}(12) = D_6$
2	2	1	$l = 6$: Exceptional group E_6 (see below)
2	1	1	$l = 5$: $\mathrm{SO}(10) = D_5$
1	1	1	$l = 4$: $\mathrm{SO}(8) = D_4$

Thus in this category we have obtained, at first in a torrent but then in a trickle, the π-systems of $\mathrm{SO}(2l) = D_l$ for $l \geq 4$; and then the three exceptional groups E_6, E_7, E_8.

We can collect and display all allowed Dynkin diagrams, hence also all possible CSLA's, in one master table. In the first three columns we list the number of branch points, double lines, triple lines, and in the last column give the diagrams and the names of the groups.

Branch Points	Double Line	Triple Line	Name and Diagram	
0	0	1	G_2	
0	1	0	$B_l = \mathrm{SO}(2l+1)$, $l \geq 3$	
			$C_l = \mathrm{USp}(2l)$, $l \geq 2$	
			F_4	
.....	.....	.....	.....	
1	0	0	$D_l = \mathrm{SO}(2l)$, $l \geq 4$	
			$E_l, l = 6, 7, 8$	
.....	.....	.....	.....	
0	0	0	$A_l = \mathrm{SU}(l+1)$, $l \geq 1$	

15.3 The Exceptional Groups

In comparison with the results known to us from Chapter 14, the new possible CSLA's found now are the groups G_2, F_4, E_6, E_7, E_8 — these are in a sense the surprises. We will not enter into a detailed discussion of any of them, but only gather some information, at first in tabular form, then say something about the root system for each.

Group	Rank	Order	Dimension of smallest UIR	Close relative of
G_2	2	14	7	$B_3 = \mathrm{SO}(7)$
F_4	4	52	26	$B_4 = \mathrm{SO}(9)$
E_6	6	78	27, 27*	$A_5 = \mathrm{SU}(6)$
E_7	7	133	56	$A_7 = \mathrm{SU}(8)$
E_8	8	248	248	$D_8 = \mathrm{SO}(16)$

Their root systems happen to be the following (where $\underline{e}$'s are unit vectors in Euclidean spaces):

G_2: Start with the 6 roots of SU(3): $\pm\underline{e}_1, (1/2)(\pm\underline{e}_1 \pm \sqrt{3}\underline{e}_2)$; to these, add on

$$\pm\sqrt{3}\underline{e}_2; \frac{\sqrt{3}}{2}(\pm\sqrt{3}\underline{e}_1 \pm \underline{e}_2), \text{making 12 in all.}$$

F_4: Start with the 32 roots of SO(9): $\pm\underline{e}_a, \pm\underline{e}_a \pm \underline{e}_b, a, b = 1, \ldots, 4$: to these, add on

$$\frac{1}{2}(\pm\underline{e}_1 \pm \underline{e}_2 \pm \underline{e}_3 \pm \underline{e}_4), \text{ making 48 in all.}$$

E_6: Start with the 30 roots of SU(6) in the (Racah) form $\underline{e}_a - \underline{e}_b, a, b = 1 \ldots 6$, using unit vectors in R^6; extend R^6 to R^7 by including $\underline{e}_7$; then add on the roots

$$\pm\sqrt{2}\underline{e}_7, \pm\frac{\underline{e}_7}{\sqrt{2}} + \frac{1}{2}(\pm\underline{e}_1 \pm \underline{e}_2 \pm \underline{e}_3 \pm \underline{e}_4 \pm \underline{e}_5 \pm \underline{e}_6)$$

with three plus and three minus signs; get 72 in all.

E_7: Start with the 56 roots of SU(8) in the (Racah) form $\underline{e}_a - \underline{e}_b, a, b = 1 \ldots 8$, using unit vectors in R^8; then add on $\frac{1}{2}(\pm\underline{e}_1 \pm \underline{e}_2 \ldots \pm \underline{e}_8)$, with four + and four − signs; get 126 in all.

E_8: Start with the 112 roots of SO(16): $\pm\underline{e}_a \pm \underline{e}_b, a, b = 1 \ldots 8$; then add on $\frac{1}{2}(\pm\underline{e}_1 \pm \underline{e}_2 \ldots \pm \underline{e}_8)$, even number of plus signs; get 240 roots in all.

The distribution of simple roots of various lengths in all CSLA can be conveniently taken as follows:

Group	**Length 1**	**Length $\sqrt{2}$**	**Length $\sqrt{3}$**
$\mathrm{SU}(l+1) = A_l$	–	l	–
$\mathrm{SO}(2l+1) = B_l$	1	$l-1$	–
$\mathrm{USp}(2l) = C_l$	$l-1$	1	–
$\mathrm{SO}(2l) = D_l$	–	l	–
G_2	1	–	1
F_4	2	2	–
$E_l, l = 6, 7, 8$	–	l	–

Chapter 16

Representations of Compact Simple Lie Algebras

In earlier Chapters we have dealt with certain specific UIR's of the groups of interest to us, namely the adjoint representation, and for each of the four classical families a defining representation. Now we shall gather and describe some aspects of a generic UIR, to be denoted $\mathcal{D}$, of any CSLA. We assume $\mathcal{D}$ acts on a complex linear vector space $\mathcal{V}$ of some dimension. The aim will be to convey an overall appreciation of structure within a UIR, as well as patterns holding for the set of all UIR's of CSLA's, without entering into detailed proofs of all statements.

16.1 Weights and Multiplicities

We have some compact simple Lie group G with Lie algebra $\mathcal{L}$, spanned by the Cartan–Weyl basis H_a, E_α. Since the H_a are commuting hermitian operators on $\mathcal{V}$, let us simultaneously diagonalise them and denote a common eigenvector by $|\underline{\mu}\ldots\rangle$:

$$H_a|\underline{\mu}\ldots\rangle = \mu_a|\underline{\mu}\ldots\rangle, a = 1, 2, \ldots, l \tag{16.1}$$

We assume $|\underline{\mu}\ldots\rangle$ is normalised to unity. Here $\underline{\mu}$ is a real l-component vector with components μ_a being the eigenvalues of H_a. We have already introduced the term *weight vector* for $\underline{\mu}$: it is a vector in the same l-dimensional Euclidean space in which roots lie. We can construct a basis for $\mathcal{V}$ with vectors $|\underline{\mu}\ldots\rangle$. Vectors $|\underline{\mu}\ldots\rangle$ with different weights $\underline{\mu}$ are clearly linearly independent.

While the H_a do commute pairwise, they may not be a complete commuting set. This may even depend on the particular UIR $\mathcal{D}$. So we have put dots after $\underline{\mu}$ in the ket vectors $|\underline{\mu}\ldots\rangle \in \mathcal{V}$: for a given $\underline{\mu}$, there may be several independent simultaneous eigenkets. This is the problem of *multiplicity*. A nondegenerate $\underline{\mu}$, i.e., a weight for which there is just one eigenket $|\underline{\mu}\rangle$, is said to be a *simple* weight (of course for the UIR $\mathcal{D}$ under consideration). Thus, one of Cartan's

theorems can be expressed as saying that the non-zero weights of the adjoint representation are all simple.

Given the UIR $\mathcal{D}$ acting on $\mathcal{V}$, we will often denote the set of all weights occurring by $\mathcal{W}$: here no attention is paid to the possible multiplicity of eigenvectors in $\mathcal{V}$ for each $\underline{\mu} \in \mathcal{W}$.

16.2 Actions of E_α and $SU(2)^{(\alpha)}$ — the Weyl Group

How do the nonhermitian generators E_α affect a ket $|\underline{\mu}\ldots\rangle$? The $H_a - E_\alpha$ commutation rule tells us that in general E_α changes the weight in a definite way — it shifts it by a definite amount:

$$\begin{aligned} H_a E_\alpha |\underline{\mu}\ldots\rangle &= [H_a, E_\alpha]|\underline{\mu}\ldots\rangle + E_\alpha H_a |\underline{\mu}\ldots\rangle \\ &= (\underline{\mu} + \underline{\alpha})_a E_\alpha |\underline{\mu}\ldots\rangle \end{aligned} \tag{16.2}$$

Therefore either $E_\alpha|\underline{\mu}\ldots\rangle$ vanishes, or it is another vector in $\mathcal{V}$ with weight $\underline{\mu} + \underline{\alpha}$. In the latter case, the multiplicities of $\underline{\mu}$ and $\underline{\mu} + \underline{\alpha}$ could be different. We write symbolically

$$E_\alpha|\underline{\mu}\ldots\rangle = (\ldots)|\underline{\mu} + \underline{\alpha}\ldots\rangle. \tag{16.3}$$

Thus the relationship (14.1) is easily understood.

Now take a root $\underline{\alpha} \in \mathfrak{R}$, its associated group $SU(2)^{(\alpha)}$, and analyse a weight $\underline{\mu} \in \mathcal{W}$ with respect to it. Recall that the $SU(2)^{(\alpha)}$ generators are

$$J_3 = \frac{\underline{\alpha} \cdot \underline{H}}{|\underline{\alpha}|^2}, J_\pm = \frac{\sqrt{2}}{|\underline{\alpha}|} E_{\pm\alpha}. \tag{16.4}$$

The ket $|\underline{\mu}\ldots\rangle$ is an eigenket of J_3:

$$\begin{aligned} J_3|\underline{\mu}\ldots\rangle &= m|\underline{\mu}\ldots\rangle, \\ m &= \underline{\alpha} \cdot \underline{\mu}/|\underline{\alpha}|^2 \end{aligned} \tag{16.5}$$

From our knowledge of UIR's of SU(2), we see that m is quantized and must have one of the values $0, \pm 1/2, \pm 1, \ldots$. More precisely,

$$\underline{\alpha} \in \mathfrak{R}, \underline{\mu} \in \mathcal{W} \Rightarrow 2\underline{\alpha} \cdot \underline{\mu}/|\underline{\alpha}|^2 = 0, \pm 1, \pm 2, \ldots \tag{16.6}$$

Again from the structure of SU(2) UIR's, we can see by applying $E_{\pm\alpha}$ repeatedly to $|\underline{\mu}\ldots\rangle$ that there must be integers $p, q \geq 0$ such that

$$\begin{aligned} &\underline{\mu} + p\underline{\alpha},\ \underline{\mu} + (p-1)\underline{\alpha},\ \ldots,\ \underline{\mu} + \underline{\alpha},\ \underline{\mu},\ \underline{\mu} - \underline{\alpha},\ \ldots,\ \underline{\mu} - (q-1)\underline{\alpha}\underline{\mu} - q\underline{\alpha} \in \mathcal{W}, \\ &\underline{\mu} + (p+1)\underline{\alpha},\ \underline{\mu} - (q+1)\underline{\alpha} \notin \mathcal{W}. \end{aligned} \tag{16.7}$$

(You can see the similarity to the root analysis in Chapter 12). In particular, there can be no gaps in this sequence of weights in $\mathcal{W}$. Now we must consider the question of multiplicities.

Consider the set of ket vectors in $\mathcal{V}$ with the above string of weights, each occurring a certain number of times:

$$|\underline{\mu}+p\underline{\alpha},\ldots\rangle,\ |\underline{\mu}+(p-1)\underline{\alpha},\ldots\rangle,\ \ldots,\ |\underline{\mu}+\underline{\alpha},\ldots\rangle,$$
$$|\underline{\mu},\ldots\rangle,\ |\underline{\mu}-\underline{\alpha},\ldots\rangle,\ \ldots,\ |\underline{\mu}-q\underline{\alpha},\ldots\rangle \tag{16.8}$$

They span some subspace of $\mathcal{V}$ which is clearly invariant under action by $SU(2)^{(\alpha)}$. So there must be present here a certain set of UIR's of $SU(2)^{(\alpha)}$ (i.e., j – values), each occurring a certain number of times. The string of J_3 eigenvalues involved is evidently,

$$m+p,\ m+p-1,\ \ldots,\ m+1,\ m,\ m-1,\ \ldots,\ m-q, \tag{16.9}$$

where m was determined in Eq.(16.5). The fact that the chain (16.8) terminates at the indicated ends means that

$$E_\alpha|\underline{\mu}+p\underline{\alpha},\ldots\rangle = E_{-\alpha}|\underline{\mu}-q\underline{\alpha},\ldots\rangle = 0 \tag{16.10}$$

Evidently in the spectrum of j-values present in the $SU(2)^{(\alpha)}$ UIR on this string of states, we have

$$j_{\max} = m+p = -(m-q),$$
$$j_{\max} = (1/2)(q+p), \quad m = (1/2)(q-p) \tag{16.11}$$

The UIR with $j = j_{\max}$ is present as often as the multiplicity of $|\underline{\mu}+p\underline{\alpha}\ldots\rangle$, which must be the same as for $|\underline{\mu}-q\underline{\alpha}\ldots\rangle$,. The UIR with $j = j_{\max}-1$ is present as often as the difference of the multiplicities of $|\underline{\mu}+(p-1)\underline{\alpha}\ldots\rangle$, and $|\underline{\mu}+p\underline{\alpha}\ldots\rangle$, and so on. [Throughout we are considering $SU(2)^{(\alpha)}$ action on just the set of vectors (16.8)]. If we wish, we could arrange to diagonalise $E_{-\alpha}E_\alpha$ on these states, which will then explicitly reveal the spectrum and multiplicity of j-values.

A characteristic feature of UIR's of SU(2) is that if the eigenvalue m of J_3 occurs, then so does $-m$; in fact within SU(2) representations, the corresponding vectors are related by a 180° rotation about the y-axis:

$$e^{i\pi J_2}|j,m\rangle \sim |j,-m\rangle \tag{16.12}$$

Applying this to the present case of $SU(2)^{(\alpha)}$, we see that if $\underline{\mu}\in\mathcal{W}$ and $\underline{\alpha}\in\mathfrak{R}$, then

$$\underline{\mu}-2\underline{\alpha}\,\underline{\alpha}\cdot\underline{\mu}/|\underline{\alpha}|^2 \in \mathcal{W} \tag{16.13}$$

with the *same* multiplicity as for $\underline{\mu}$. This is because

$$\exp(i\pi J_2^{(\alpha)})(\text{set of states } |\underline{\mu}\ldots\rangle) = \text{set of states } |\underline{\mu}-2\underline{\alpha}\,\underline{\alpha}\cdot\underline{\mu}/|\underline{\alpha}|^2,\ldots\rangle \tag{16.14}$$

and $\exp(i\pi J_2^{(\alpha)})$ is evidently unitary.

We can summarise these results of applying $SU(2)^{(\alpha)}$ to $\underline{\mu} \in \mathcal{W}$:

(i) there exist integers $p, q \geq 0$ such that Eq.(16.7) holds;

(ii) $2\underline{\mu} \cdot \underline{\alpha}/|\underline{\alpha}|^2 = q - p = 0, \pm 1, \pm 2, \ldots$;

(iii) $\underline{\mu} - 2\underline{\alpha}\,\underline{\alpha} \cdot \underline{\mu}/|\underline{\alpha}|^2 \in \mathcal{W}$, with the same multiplicity as $\underline{\mu}$.

The reflection operation $\exp(i\pi J_2^{(\alpha)})$ is an important and useful one. For each $\underline{\alpha} \in \mathfrak{R}$, one such reflection is defined. In weight space, this is clearly reflection through a plane perpendicular to $\underline{\alpha}$ passing through the origin. The set of all such reflections is a group, the Weyl group; this is a finite subgroup of G. If $\underline{\mu} \in \mathcal{W}$, any element of the Weyl group applied to $\underline{\mu}$ gives some $\underline{\mu}' \in \mathcal{W}$ with the same multiplicity as $\underline{\mu}$. We say $\underline{\mu}, \underline{\mu}' \in \mathcal{W}$ are *equivalent*, if they are related by an element of the Weyl group. So equivalent weights have the same multiplicity, and

Set of weights equivalent to

$$\underline{\mu} \in \mathcal{W} = \{\underline{\mu} - 2\underline{\alpha}\,\underline{\alpha} \cdot \underline{\mu}/|\underline{\alpha}|^2 \,|\, \underline{\alpha} \in \mathfrak{R}\}. \tag{16.15}$$

16.3 Dominant Weights, Highest Weight of a UIR

We are working with a definite choice and sequence of the operators H_a in the Cartan subalgebra. We then say: a weight $\underline{\mu}$ is *higher* than a weight $\underline{\mu}'$ if in the difference $\underline{\mu} - \underline{\mu}'$ (which may not be a weight!) the first non-zero component is *positive*. So for any two distinct weights, one is definitely higher than the other.

The highest weight in a set of equivalent weights, i.e., a set related by Weyl reflections, is called a *dominant* weight.

Given the UIR $\mathcal{D}$ acting on $\mathcal{V}$, and the set $\mathcal{W}$ of all weights that occur, denote by $\underline{\Lambda} \in \mathcal{W}$ the highest weight. Then it is a simple weight with no multiplicity or degeneracy; there is just one ket $|\underline{\Lambda}\rangle$. We indicate the proof later on. Clearly, if $\underline{\Lambda}$ is the highest weight, it means that if $\underline{\alpha}$, is any positive root, $\underline{\Lambda} + \underline{\alpha}$ is not present in $\mathcal{W}$:

$$\underline{\alpha} \in \mathfrak{R}_+ \Rightarrow \underline{\Lambda} + \underline{\alpha} \notin \mathcal{W}, E_\alpha|\underline{\Lambda}\rangle = 0 \tag{16.16}$$

Even more economically stated, this is the same as

$$\underline{\alpha}^{(a)} \in \mathcal{S}: \quad E_{\alpha^{(a)}}|\underline{\Lambda}\rangle = 0, a = 1, 2, \ldots, l \tag{16.17}$$

So for a general compact simple Lie group G, the set of l "raising operators" $E_{\alpha^{(a)}}$ for $\underline{\alpha}^{(a)} \in \mathcal{S}$ generalises the single J_+ of SU(2). The highest weight state $|\underline{\Lambda}\rangle$ in $\mathcal{V}$ is simultaneously the maximum magnetic quantum number state $(m = j)$ for the $SU(2)^{(\alpha)}$ algebras for all $\underline{\alpha} \in \mathfrak{R}_+$. This means:

$$\underline{\alpha} \in \mathfrak{R}_+: \quad 2\underline{\Lambda} \cdot \underline{\alpha}/|\underline{\alpha}|^2 = \text{integer} \geq 0 \tag{16.18}$$

We shall introduce the notation

$$\underline{\alpha}^{(a)} \in S: \quad 2\underline{\Lambda} \cdot \underline{\alpha}^{(a)}/|\underline{\alpha}^{(a)}|^2 = N_a \geq 0, \qquad \underline{\Lambda} \longleftrightarrow \{N_a\} \tag{16.19}$$

It turns out that each UIR can be uniquely characterised by its highest weight $\underline{\Lambda}$, or by giving the nonnegative integers N_a.

Let us indicate why the highest weight state $|\underline{\Lambda}\rangle$ in a UIR is nondegenerate. Clearly we can go from $|\underline{\Lambda}\rangle$ to other (basis) kets in $\mathcal{V}$ by action with E_α for $\underline{\alpha} \in \mathfrak{R}_-$. Now suppose there are two highest weight states $|\underline{\Lambda}, 1\rangle$ and $|\underline{\Lambda}, 2\rangle$, mutually orthogonal. With no loss of generality:

$$\langle \underline{\Lambda}, 2|\underline{\Lambda}, 1\rangle = 0 \tag{16.20}$$

Then for any $\underline{\alpha} \in \mathfrak{R}_-$, the states $E_\alpha|\underline{\Lambda}, 1\rangle$ and $E_\alpha|\underline{\Lambda}, 2\rangle$ will also be orthogonal:

$$\begin{aligned} \underline{\alpha} \in \mathfrak{R}_- \Rightarrow -\underline{\alpha} \in \mathfrak{R}_+ : \langle \underline{\Lambda}, 2|E_\alpha^+ E_\alpha|\underline{\Lambda}, 1\rangle &= \langle \underline{\Lambda}, 2|E_{-\alpha} E_\alpha|\underline{\Lambda}, 1\rangle \\ &= \langle \underline{\Lambda}, 2|[E_{-\alpha}, E_\alpha]|\underline{\Lambda}, 1\rangle \\ &= -\underline{\alpha} \cdot \underline{\Lambda}\langle \underline{\Lambda}, 2|\underline{\Lambda}, 1\rangle = 0 \end{aligned} \tag{16.21}$$

One can continue this argument for chains like $E_\beta E_\alpha|\underline{\Lambda}, 1 \text{ or } 2\rangle$, and finds similar results. So if the highest weight state has degeneracy, the whole space $\mathcal{V}$ splits into two or more orthogonal subspaces, unconnected by the generators, hence the representation must be reducible. Conversely, in a UIR the highest weight is multiplicity free. It also is true that two UIR's with the same highest weight are equivalent.

Now let us return to the problem of classifying UIR's. It turns out that there is one UIR (up to unitary equivalence) for each given set of nonnegative integers $\{N_a\}$ defined in Eq.(16.19), and conversely:

$$\begin{aligned} &N_a \geq 0, a = 1, 2, \ldots, l \leftrightarrow \text{unique UIR } \{N_a\} \text{ of } G, \text{highest weight } \underline{\Lambda}: \\ &\quad 2\underline{\Lambda} \cdot \underline{\alpha}^{(a)}/|\underline{\alpha}^{(a)}|^2 = N_a \end{aligned} \tag{16.22}$$

Can we explicitly exhibit this $\underline{\Lambda}$ once $\{N_a\}$ are given? We simply use the fact that the set of simple roots $S = \{\underline{\alpha}^{(a)}\}$ is linearly independent and forms a basis for root and weight space. The "problem" is that we have an oblique non-orthonormal basis. Nevertheless, linear independence of the $\underline{\alpha}^{(a)}$ allows us to expand $\underline{\Lambda}$ as

$$\underline{\Lambda} = \sum_{a=1}^{l} \lambda_a \underline{\alpha}^{(a)}, \lambda_a \text{ real} \tag{16.23}$$

The linear relations between λ_a and N_a then are

$$N_a = \sum_{b=1}^{l} 2(\underline{\alpha}^{(b)} \cdot \underline{\alpha}^{(a)}/|\underline{\alpha}^{(a)}|^2) \cdot \lambda_b \tag{16.24}$$

This can and must be inverted. Let us define the *Cartan matrix* A, a real $l \times l$ matrix, as

$$A = (A_{ab}), A_{ab} = 2\underline{\alpha}^{(a)} \cdot \underline{\alpha}^{(b)}/|\underline{\alpha}^{(b)}|^2 \tag{16.25}$$

It has extremely simple elements. Recalling that the $\underline{\alpha}^{(a)}$ are simple roots, we have the values:

$$\begin{aligned} & A_{aa} = 2, \\ a \neq b: \ & A_{ab} = 0 \quad \text{for} \quad \underline{\alpha}^{(a)} \cdot \underline{\alpha}^{(b)} = 0, \quad \text{i.e., O} \ \ \text{O}, \\ & \quad = -1 \text{ for} \qquad\quad \text{O–O} \\ & \quad = -1 \text{ or } -2 \text{ for O=O} \\ & \quad = -1 \text{ or } -3 \text{ for O}\equiv\text{O} \end{aligned} \tag{16.26}$$

Even though this matrix is not symmetric, the simple values of its entries make it easy to work with. If we wish, we can express A as the product of a symmetric and a diagonal matrix:

$$\begin{aligned} A &= SD, \\ S_{ab} &= S_{ba} = \underline{\alpha}^{(a)} \cdot \underline{\alpha}^{(b)}, \\ D_{ab} &= 2\delta_{ab}/|\underline{\alpha}^{(a)}|^2 \end{aligned} \tag{16.27}$$

Linear independence of the simple roots means that A is nonsingular, but A^{-1} is again not symmetric. On the other hand, the matrix

$$G = A^{-1}D^{-1} = D^{-1}S^{-1}D^{-1} \tag{16.28}$$

is symmetric. We shall express the solution for the highest weight $\underline{\Lambda}$ in terms of the set of integers $\{N_a\}$ by using this matrix G. The equation to be inverted, Eq.(16.24), reads in matrix form:

$$N = A^T\lambda \tag{16.29}$$

so the solution is

$$\lambda = (A^T)^{-1}N = (GD)^T N = DGN \tag{16.30}$$

Writing out the individual components and putting them into Eq.(16.23),

$$\begin{aligned} \lambda_a &= \frac{2}{|\underline{\alpha}^{(a)}|^2} \sum_{b=1}^{l} G_{ab}N_b \\ \underline{\Lambda} &= \sum_{a,b} 2G_{ab}N_b \, \underline{\alpha}^{(a)}/|\underline{\alpha}^{(a)}|^2 \end{aligned} \tag{16.31}$$

This way of writing $\underline{\Lambda}$ in terms of $\{N_a\}$ leads to the next important step. We said that any choice of the N's leads to a unique UIR and vice versa. We can then consider the l special UIR's obtained by setting all N_a but one zero,

and taking the non-zero one to be unity. In this way certain special highest weights must arise. We write them as

$$\underline{\Lambda}^{(a)} = \sum_b 2G_{ab}\,\underline{\alpha}^{(b)}/|\underline{\alpha}^{(b)}|^2$$
$$\underline{\Lambda} = \sum_a N_a \underline{\Lambda}^{(a)} \tag{16.32}$$

We use these particular highest weights in the following section.

16.4 Fundamental UIR's, Survey of all UIR's

Cartan has shown that for any CSLA, there are l basic or fundamental UIR's out of which all others can be built via symmetrised direct products. We denote the fundamental UIR's by $\mathcal{D}^{(a)}, a = 1, 2, \ldots, l$:

$$\begin{aligned}\mathcal{D}^{(a)} &= a^{\text{th}} \text{ fundamental UIR with highest weight } \underline{\Lambda}^{(a)} \\ &= \text{UIR with } N_a = 1, N_b = 0 \text{ for } b \neq a\end{aligned} \tag{16.33}$$

In this sense, these $\underline{\Lambda}^{(a)}$ are the simplest highest weights: they are called *fundamental dominant weights.* Any dominant weight (i.e., the highest in an equivalence class under the Weyl group) is a non-negative integral combination of the $\underline{\Lambda}^{(a)}$. The highest weight of a UIR is the linear combination (16.32).

The UIR uniquely determined by $\{N_a\}$ can be indicated by use of the Dynkin diagram for the group. On this diagram we mark the positions of the first, second, ..., l^{th} simple roots $\underline{\alpha}^{(1)}, \underline{\alpha}^{(2)}, \ldots \underline{\alpha}^{(l)}$. Then just below or at the corresponding circles we write the values of $N_1, N_2, \ldots, N_l$. In this depiction, $\mathcal{D}^{(a)}$ has unity at the a^{th} root position, zero elsewhere.

How is the UIR $\mathcal{D} \equiv \{N_a\}$ obtained from the fundamental UIR's $\mathcal{D}^{(a)}$? The prescription is to take N_1 products of $\mathcal{D}^{(1)}, N_2$ of $\mathcal{D}^{(2)}, \ldots, N_l$ of $\mathcal{D}^{(l)}$, take the direct product of them all, and isolate the highest weight in this product. This highest weight is evidently given by Eq.(16.32). Thus on reduction of this direct product representation, the "largest piece" is the UIR $\{N_a\}$ we are after:

$$\begin{aligned}&\mathcal{D}^{(1)} \times \ldots \times \mathcal{D}^{(1)} \times \mathcal{D}^{(2)} \times \ldots \mathcal{D}^{(2)} \times \ldots \mathcal{D}^{(l)} \times \ldots \mathcal{D}^{(l)} \rightarrow \\ &\longleftarrow\!\!N_1\!\!\longrightarrow \qquad \longleftarrow N_2 \longrightarrow \qquad \longleftarrow N_l \longrightarrow \\ &\underline{\Lambda} = N_1\underline{\Lambda}^{(1)} + N_2\underline{\Lambda}^{(2)} + \ldots N_l\underline{\Lambda}^{(l)} - \text{UIR}\{N_a\}\end{aligned} \tag{16.34}$$

The fundamental dominant weights $\underline{\Lambda}^{(a)}$ determining the fundamental UIR's out of which all UIR's can be built, have a simple geometrical relationship to the simple roots $\mathcal{S}$. Since each $\underline{\Lambda}^{(a)}$ corresponds to $N_a = 1, N_b = 0$ for $b \neq a$, we see that

$$2\underline{\Lambda}^{(a)} \cdot \underline{\alpha}^{(b)}/|\underline{\alpha}^{(b)}|^2 = \delta_{ab} \tag{16.35}$$

We therefore set up the following procedure. Starting with the simple roots $\underline{\alpha}^{(a)}$, we rescale them to define the linearly independent (but not normalised!)

vectors

$$\hat{\alpha}^{(a)} = 2\underline{\alpha}^{(a)}/|\underline{\alpha}^{(a)}|^2 \tag{16.36}$$

Then the fundamental dominant weights form the *reciprocal basis* in root space to this set of vectors

$$\underline{\Lambda}^{(a)} \cdot \hat{\alpha}^{(b)} = \delta_{ab} \tag{16.37}$$

Incidentally, these vectors $\hat{\alpha}^{(a)}$ are related to the symmetric matrix G appearing in Eqs.(16.28),(16.30),(16.31),(16.32):

$$G^{-1} = DSD = (\hat{\alpha}^{(a)} \cdot \hat{\alpha}^{(b)});$$
$$G = (\hat{\alpha}^{(a)} \cdot \hat{\alpha}^{(b)})^{-1} \tag{16.38}$$

The integers $\{N_a\}$ label UIR's uniquely, and they tell us graphically how to build up general UIR's from the fundamental ones. Are there invariant operators such that the N_a or functions thereof are eigenvalues of these operators? There are, namely the Casimir operators which are (symmetric) polynomials formed out of the generators X_j. For example, the simplest quadratic one is

$$\mathcal{C}_2 = g^{jk} X_j X_k$$
$$= \sum_{a=1}^{l} H_a^2 + \sum_{\alpha \in \mathfrak{R}} E_{-\alpha} E_\alpha \tag{16.39}$$

Then there are cubic and higher order expressions, in fact precisely l independent Casimir operators. However we will not pursue their properties in any detail.

For a CSLA of rank l, when is the UIR $\{N_a\}$ real or potentially real? When is it essentially complex? Just as a UIR has a unique simple highest weight $\underline{\Lambda}$, it also has a lowest simple weight $\underline{\nu}$. Then the UIR $\{N_a\}$ is equivalent to its complex conjugate if and only if $\underline{\Lambda} = -\underline{\nu}$. More generally, if $\underline{\Lambda}$ and $\underline{\nu}$ are the highest and lowest weights of the UIR $\mathcal{D}$, then $-\underline{\nu}$ and $-\underline{\Lambda}$ are the highest and lowest weights of the UIR $\mathcal{D}^*$.

From this general criterion we see immediately how it is that each UIR of SU(2) is self conjugate: the maximum and minimum values of m, j and $-j$, are negatives of one another.

We have seen that the complete list of compact simple Lie groups is SU(n) for $n \geq 2$; SO(n) for $n \geq 7$; Usp($2n$) for $n \geq 2$; G_2; F_4; E_n for $n = 6, 7, 8$. Which of these possess complex UIR's? A detailed analysis shows that only SU(n) for $n \geq 3$, SO($4n+2$) for $n \geq 2$ and E_6 have some complex UIR's; in all other cases each UIR is real or pseudo real.

16.5 Fundamental UIR's for A_l, B_l, C_l, D_l

The simple roots for each of these groups have been given in Chapter 14. In principle we can then calculate the fundamental dominant weights as the basis reciprocal to $\{\hat{\alpha}^{(a)}\}$. The calculations are quite easy for B_l, C_l and D_l, and a bit involved for A_l. We will take them up in the same sequence D_l, B_l, C_l, A_l in which we dealt with them in Chapter 14.

Case of $D_l \equiv \mathbf{SO}(2l)$:

With $\underline{e}_a$ being the standard unit vectors in Euclidean l-dimensional space, we have from Eq.(14.20) the following set of simple roots:

$$\underline{\alpha}^{(1)} = \underline{e}_1 - \underline{e}_2,\ \underline{\alpha}^{(2)} = \underline{e}_2 - \underline{e}_3,\ \ldots,\ \underline{\alpha}^{(l-1)} = \underline{e}_{l-1} - \underline{e}_l,\ \underline{\alpha}^{(l)} = \underline{e}_{l-1} + \underline{e}_l \quad (16.40)$$

Since each $|\underline{\alpha}^{(a)}| = \sqrt{2}$, the rescaling (16.36) has no effect:

$$\underline{\alpha}^{(a)} = \hat{\alpha}^{(a)} \quad (16.41)$$

So the fundamental dominant weights $\underline{\Lambda}^{(1)}, \underline{\Lambda}^{(2)}, \ldots$ form a system reciprocal to the set (16.40). At this point let us remember from Eq.(14.13) that in the defining vector representation D of $\mathrm{SO}(2l)$, the highest weight is $\underline{e}_1$. In fact, the set of all weights in this representation can be arranged in decreasing order thus:

$$\begin{aligned} \mathcal{W}_D &= \text{ weights in defining representation } D \text{ of } \mathrm{SO}(2l) \\ &= \{\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_l, -\underline{e}_l, \ldots, -\underline{e}_2, -\underline{e}_1\} \end{aligned} \quad (16.42)$$

We will use this information in interpreting the fundamental representations.

Let us now list the fundamental dominant weights $\underline{\Lambda}^{(a)}$; they are easy to find from the condition that they form a basis reciprocal to the simple roots (16.40):

$$\begin{aligned} \underline{\Lambda}^{(1)} &= \underline{e}_1,\ \ \underline{\Lambda}^{(2)} = \underline{e}_1 + \underline{e}_2,\ \ \underline{\Lambda}^{(3)} = \underline{e}_1 + \underline{e}_2 + \underline{e}_3, \ldots, \\ \underline{\Lambda}^{(l-2)} &= \underline{e}_1 + \underline{e}_2 + \ldots + \underline{e}_{l-2}, \\ \underline{\Lambda}^{(l-1)} &= \frac{1}{2}(\underline{e}_1 + \underline{e}_2 + \ldots + \underline{e}_{l-1} - \underline{e}_l), \\ \underline{\Lambda}^{(l)} &= \frac{1}{2}(\underline{e}_1 + \underline{e}_2 + \ldots + \underline{e}_{l-1} + \underline{e}_l) \end{aligned} \quad (16.43)$$

The "half integral" structure of the last two must come as somewhat of a surprise! What can we say about the corresponding fundamental UIR's $\mathcal{D}^{(1)}, \mathcal{D}^{(2)}, \ldots, \mathcal{D}^{(l)}$? Comparing with the weights $\mathcal{W}_D$ of the defining representation D in Eq.(16.42), we can easily see that

$$\begin{aligned} \mathcal{D}^{(1)} &= \text{defining } \textit{vector} \text{ representation } D \\ \mathcal{D}^{(2)} &= \text{representation on second rank antisymmetric} \\ &\quad \text{tensors, } (D \times D)_{\text{antisymmetric}} \\ \mathcal{D}^{(3)} &= \text{representation on third rank antisymmetric} \\ &\quad \text{tensors, } (D \times D \times D)_{\text{antisymmetric}} \\ &\quad \ldots\ldots \\ \mathcal{D}^{(l-2)} &= \text{representation on } (l-2) \text{ rank antisymmetric} \\ &\quad \text{tensors.} \end{aligned} \quad (16.44)$$

The truth of these statements would be evident to you by using arguments similar to those in quantum mechanics when enumerating the possible states of identical fermions obeying the exclusion principle. So all the fundamental UIR's *except the last two* really arise out of the defining representation D, and its products with itself. The last two, however, can *never* be obtained by any finite algebraic procedure starting with D; this is evident on comparing $\underline{\Lambda}^{(l-1)}$ and $\underline{\Lambda}^{(l)}$with $\mathcal{W}_D$! These last two "unexpected" fundamental UIR's of SO($2l$) are called *spinor representations.*

$$\begin{aligned} \mathcal{D}^{(l-1)} &= \text{first spinor representation, dimension } 2^{l-1}, \text{ also} \\ &\quad \text{written } \Delta^{(1)}; \\ \mathcal{D}^{(l)} &= \text{second spinor representation, dimension } 2^{l-1}, \\ &\quad \text{also written } \Delta^{(2)} \end{aligned} \tag{16.45}$$

We shall study the construction and properties of these spinor UIR's in the next Chapter. At this point, it suffices to repeat that the l fundamental UIR's of SO($2l$) are the antisymmetric tensor representations of ranks $1, 2, \ldots, l-2$, and the two (inequivalent) spinor representations $\Delta^{(1)}, \Delta^{(2)}$.

Associated with the group SO($2l$) is an antisymmetric Levi-Civita symbol carrying $2l$ indices. So one knows that among antisymmetric tensors there is no need to go beyond those of rank l. One naturally asks: where are the antisymmetric tensors of ranks $l-1$ and l, and what status do they have? They are of course not in the list of fundamental UIR's. It turns out that the $(l-1)^{\text{th}}$ rank antisymmetric tensors furnish a UIR with the highest weight

$$\begin{aligned} \underline{\Lambda} &= \underline{e}_l + \underline{e}_2 + \ldots + \underline{e}_{l-1} \\ &= \underline{\Lambda}^{(l-1)} + \underline{\Lambda}^{(l)} \end{aligned} \tag{16.46}$$

Thus, in the $\{N_a\}$ notation, this is the UIR $(0, 0, \ldots, 1, 1)$, and it occurs in the direct product $\Delta^{(1)} \times \Delta^{(2)}$ of the two spinor UIR's. As for the l^{th} rank antisymmetric tensors, they are reducible into self dual and antiself dual types, and depending on one's definitions, one is present in $\{\Delta^{(1)} \times \Delta^{(1)}\}$ symm and the other in $\{\Delta^{(2)} \times \Delta^{(2)}\}_{\text{symm}}$. We shall deal with these in some detail in the next Chapter.

Case of $B_l = \mathbf{SO}(2l+1)$:

The set of simple roots in this case is, from Eq.(14.28),

$$\underline{\alpha}^{(1)} = \underline{e}_1 - \underline{e}_2,\ \underline{\alpha}^{(2)} = \underline{e}_2 - \underline{e}_3,\ \ldots,\ \underline{\alpha}^{(l-1)} = \underline{e}_{l-l} - \underline{e}_l,\ \underline{\alpha}^{(l)} = \underline{e}_l \tag{16.47}$$

After rescaling according to Eq.(16.36), we get the vectors $\hat{\alpha}^{(a)}$:

$$\hat{\alpha}^{(1)} = \underline{e}_1 - \underline{e}_2,\ \hat{\alpha}^{(2)} = \underline{e}_2 - \underline{e}_3,\ \ldots,\ \hat{\alpha}^{(l-1)} = \underline{e}_{l-1} - \underline{e}_l,\ \hat{\alpha}^{(l)} = 2\underline{e}_l \tag{16.48}$$

Now the set of weights in the defining representation, arranged from the highest down to the lowest, are:

$$\begin{aligned} \mathcal{W}_D &= \text{weights in defining representation } D \text{ of SO}(2l+1) \\ &= \{\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_l, \underline{0}, -\underline{e}_l, \ldots, -\underline{e}_2, -\underline{e}_1\} \end{aligned} \tag{16.49}$$

We now have the requisite information to construct the fundamental dominant weights and interpret the corresponding fundamental UIR's: the former are

$$\begin{aligned} &\underline{\Lambda}^{(1)} = \underline{e}_1, \quad \underline{\Lambda}^{(2)} = \underline{e}_1 + \underline{e}_2, \quad \underline{\Lambda}^{(3)} = \underline{e}_1 + \underline{e}_2 + \underline{e}_3, \ldots, \\ &\underline{\Lambda}^{(l-1)} = \underline{e}_1 + \underline{e}_2 + \ldots + \underline{e}_{l-1} \\ &\underline{\Lambda}^{(l)} = \frac{1}{2}(\underline{e}_1 + \underline{e}_2 + \ldots + \underline{e}_l) \end{aligned} \tag{16.50}$$

We are thus able to recognise the first $(l-1)$ fundamental UIR's:

$$\begin{aligned} \mathcal{D}^{(1)} &= \text{defining } \textit{vector} \text{ representation } D \\ \mathcal{D}^{(2)} &= \text{representation on second rank antisymmetric} \\ &\quad \text{tensors, } (D \times D)_{\text{antisymmetric}} \\ &\quad \ldots\ldots \\ \mathcal{D}^{(l-1)} &= \text{representation on } (l-1)^{\text{th}} \text{ rank antisymmetric} \\ &\quad \text{tensors} \end{aligned} \tag{16.51}$$

In contrast to SO($2l$), at the end of this list there is now only *one* spinor UIR:

$$\begin{aligned} \mathcal{D}^{(l)} &= \text{unique spinor representation, dimension } 2^l, \\ &\quad \text{written } \Delta \end{aligned} \tag{16.52}$$

This spinor representation can, of course, not be obtained from D by any finite algebraic procedure. We study it in detail in Chapter 17.

For SO($2l+1$), the independent antisymmetric tensors are those of ranks $1, 2, \ldots, l-1, l$. All but the last have appeared in the list of fundamental UIR's. The rank l antisymmetric tensors do not yield a fundamental UIR. The highest weight for their representation is clearly the sum of the first l weights in D listed in Eq.(16.49):

$$\underline{\Lambda} = \underline{e}_1 + \underline{e}_2 + \ldots + \underline{e}_l = 2\underline{\Lambda}^{(l)} \tag{16.53}$$

Therefore, in the $\{N_a\}$ notation they furnish the UIR $(0, 0, \ldots, 0, 2)$, and this is the leading UIR in the symmetric part of the product $\Delta \times \Delta$.

Case of $C_l = \mathbf{USp}(2l)$

Interestingly, the simple roots and the scaled simple roots now are just interchanged as compared to B_l =SO$(2l+1)$: from Eq.(14.48),

$$\underline{\alpha}^{(1)} = \underline{e}_1 - \underline{e}_2,\ \underline{\alpha}^{(2)} = \underline{e}_2 - \underline{e}_3,\ \ldots,\ \underline{\alpha}^{(l-1)} = \underline{e}_{l-1} - \underline{e}_l,\ \underline{\alpha}^{(l)} = 2\underline{e}_l;$$
$$\hat{\alpha}^{(1)} = \underline{e}_1 - \underline{e}_2,\ \hat{\alpha}^{(2)} = \underline{e}_2 - \underline{e}_3,\ \ldots,\ \hat{\alpha}^{(l-1)} = \underline{e}_{l-1} - \underline{e}_l,\ \hat{\alpha}^{(l)} = \underline{e}_l \qquad (16.54)$$

The weights of the defining representation given in Eq.(14.43) are the same as with SO$(2l)$:

$$\mathcal{W}_D = \{\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_l, -\underline{e}_l, -\underline{e}_{l-1}, \ldots, -\underline{e}_1\} \qquad (16.55)$$

The reciprocal basis to $\hat{\alpha}^{(a)}$ is very easily found. There are no half integers anywhere now:

$$\underline{\Lambda}^{(1)} = \underline{e}_1, \quad \underline{\Lambda}^{(2)} = \underline{e}_1 + \underline{e}_2 + \cdots,$$
$$\underline{\Lambda}^{(l-1)} = \underline{e}_1 + \underline{e}_2 + \cdots + \underline{e}_{l-1},$$
$$\underline{\Lambda}^{(l)} = \underline{e}_1 + \underline{e}_2 + \cdots + \underline{e}_l \qquad (16.56)$$

There is a uniform pattern now all the way up to the last one. We can say that the fundamental UIR's $\mathcal{D}^{(a)}, a = 1, 2, \ldots, l$, are given by the ath rank antisymmetric tensors over the defining vector UIR D, except that in this symplectic case it is the antisymmetric objects that must have traces removed! The trace operation of course involves contraction with the symplectic metric η_{AB} of Eq.(14.30). So the fundamental UIR's for USp$(2l)$ are

$$a = 1, 2, \ldots, l: \quad \mathcal{D}^{(a)} = a^{\text{th}} \text{ rank antisymmetric traceless tensors over defining vector representation } D \qquad (16.57)$$

There are no spinors for the symplectic groups.

Case of $A_l \equiv \mathbf{SU}(l+1)$

We have noticed in Chapter 14 that the calculations of weights and roots for the groups SU$(l+1)$ tend to be slightly messy compared to the other families! The simple roots are listed in Eq.(14.65). Each of them has length $\sqrt{2}$, so we have no change on rescaling. We write them a bit more compactly as

$$\underline{\alpha}^{(a)} = \hat{\alpha}^{(a)} = \frac{1}{\sqrt{a+1}}(\sqrt{a}\underline{e}_a - \sqrt{a+2}\underline{e}_{a+1}), a = 1, 2, \ldots, l-1,$$
$$\underline{\alpha}^{(l)} = \hat{\alpha}^{(l)} = \sum_{a=1}^{l-1} \frac{\underline{e}_a}{\sqrt{a(a+1)}} + \sqrt{\frac{l+1}{l}}\underline{e}_l \qquad (16.58)$$

Next, in listing the weights of the defining representation D, given in Eq.(14.56), it is helpful to define the following unnormalised multiples of the unit vectors $\underline{e}_a$:

$$\underline{f}_a = \underline{e}_a/\sqrt{a(a+1)}, \quad a = 1, 2, \ldots, l \qquad (16.59)$$

Then the highest weight in D, followed by the successively lower weights are:

$$\begin{aligned}
\underline{\mu}^{(1)} &= \underline{f}_1 + \underline{f}_2 + \cdots + \underline{f}_l; \\
\underline{\mu}^{(l+1)} &= -l\underline{f}_l; \\
\underline{\mu}^{(l)} &= -(l-1)\underline{f}_{l-1} + \underline{f}_l; \\
\underline{\mu}^{(l-1)} &= -(l-2)\underline{f}_{l-2} + \underline{f}_{l-1} + \underline{f}_l; \\
&\cdots\cdots\cdots\cdots \\
\underline{\mu}^{(3)} &= -2\underline{f}_2 + \underline{f}_3 + \cdots + \underline{f}_{l-1} + \underline{f}_l; \\
\underline{\mu}^{(2)} &= -\underline{f}_1 + \underline{f}_2 + \cdots + \underline{f}_{l-1} + \underline{f}_l.
\end{aligned} \tag{16.60}$$

Now we must construct the l fundamental dominant weights. First we recognise the role of the defining UIR D. Its highest weight $\underline{\mu}^{(1)}$ is seen to obey

$$\begin{aligned}
&\underline{\mu}^{(1)} \cdot \hat{\alpha}^{(a)} = 0, \quad a = 1, 2, \ldots, l-1; \\
&\underline{\mu}^{(1)} \cdot \hat{\alpha}^{(l)} = 1
\end{aligned} \tag{16.61}$$

Thus according to the general definition, this is the lth fundamental UIR:

$$\underline{\mu}^{(1)} = \underline{\Lambda}^{(l)}, \quad D \equiv \mathcal{D}^{(l)} \tag{16.62}$$

To find the other fundamental dominant weights, it is helpful to expand a general vector in the form

$$\underline{x} = \sum_{a=1}^{l} x_a \underline{f}_a = \sum_{a=1}^{l} x_a \underline{e}_a / \sqrt{a(a+1)} \tag{16.63}$$

so that products with the simple roots simplify:

$$\begin{aligned}
&\underline{x} \cdot \hat{\alpha}^{(a)} = (x_a - x_{a+1})/(a+1), \quad a = 1, 2, \ldots, l-1; \\
&\underline{x} \cdot \hat{\alpha}^{(l)} = \sum_{a=1}^{l-1} x_a / a(a+1) + \frac{x_l}{l}.
\end{aligned} \tag{16.64}$$

Using such expansions, one can quickly discover the reciprocal basis to the $\hat{\alpha}^{(a)}$:

$$\begin{aligned}
\underline{\Lambda}^{(a)} &= \underline{f}_1 + \underline{f}_2 + \cdots + \underline{f}_a - a(\underline{f}_{a+1} + \ldots + \underline{f}_l) \\
&= \sum_{b=1}^{a} \underline{e}_b / \sqrt{b(b+1)} - a \sum_{b=a+1}^{l} \underline{e}_b / \sqrt{b(b+1)}, \quad a = 1, 2, \ldots, l-1
\end{aligned} \tag{16.65}$$

The lth fundamental dominant weight $\underline{\Lambda}^{(l)} = \underline{\mu}^{(1)}$ can also be taken to be given by this same general expression. These $\underline{\Lambda}^{(a)}$ can now be expressed in terms of

the successive weights of D listed in Eq.(16.60), namely, beginning with (16.62):

$$\begin{aligned}
\underline{\Lambda}^{(l)} &= \underline{\mu}^{(1)}; \\
\underline{\Lambda}^{(l-1)} &= \underline{\mu}^{(1)} + \underline{\mu}^{(l+1)}; \\
\underline{\Lambda}^{(l-2)} &= \underline{\mu}^{(1)} + \underline{\mu}^{(l+1)} + \underline{\mu}^{(l)}; \\
&\cdots\cdots\cdots\cdots \\
\underline{\Lambda}^{(1)} &= \underline{\mu}^{(1)} + \underline{\mu}^{(l+1)} + \underline{\mu}^{(l)} + \cdots + \underline{\mu}^{(3)}
\end{aligned} \tag{16.66}$$

These expressions show that the l fundamental UIR's of $A_l = \mathrm{SU}(l+1)$ are the defining representation D, and the antisymmetric tensors over D of ranks $2, 3, \ldots, l$ as one might have expected:

$$a = 1, 2, \ldots: \quad \mathcal{D}^{(a)} = \text{antisymmetric tensors of rank } (l+1-a) \text{ over } D \tag{16.67}$$

As with the symplectic groups, there are no spinors for the unitary groups.

One might be tempted to rework the algebra for the A_l family of groups to avoid the rather "peculiar" patterns for $\underline{\Lambda}^{(a)}$, weights $\underline{\mu}$ in D, etc. that we have found. However, that is bound to introduce complications elsewhere, for instance in the choice of the Cartan subalgebra generators H_a. We have made what seems like a simple and uncomplicated choice of H_a to begin with, and then gone through the general theory in a systematic way.

16.6 The Elementary UIR's

The l fundamental UIR's of a rank l group G are the building blocks for all UIR's in the sense that the general UIR $\{N_a\}$ is the "largest" piece in the direct product of N_1 factors $\mathcal{D}^{(1)}, N_2$ factors $\mathcal{D}^{(2)}, \ldots$. It is isolated by working with the highest weight in this product,

$$\underline{\Lambda}\{N\} = N_1\underline{\Lambda}^{(1)} + N_2\underline{\Lambda}^{(2)} + \ldots, \tag{16.68}$$

evidently lying in the product of the completely symmetric parts of $\{\otimes\mathcal{D}^{(1)}\}^{N_1}$, $\{\otimes\mathcal{D}^{(2)}\}^{N_2}, \ldots$.

But we have seen that for A_l, B_l, C_l and D_l, many of the fundamental UIR's are antisymmetric tensors over the defining UIR D! Thus, if we have the defining UIR in hand, and may be one or two more, then by the combined processes of antisymmetrization and symmetrization of products, all the other basic UIR's and then all UIR's can be formed. The number of elementary UIR's needed for this construction is just 2 or 3. The precise definition is that they correspond to the "ends" of the Dynkin diagram. So in the sequence D_l, B_l, C_l, A_l, we have

the elementary UIR's:

$D_l \equiv$SO$(2l)$ - Three elementary UIR's:

$\{N_a\} = \{1, 0, \ldots\} \rightarrow$ vector UIR, D;

$\{N_a\} = \{0, 0, \ldots, 1, 0\} \rightarrow$ spinor $\Delta^{(1)}$

$\{N_a\} = \{0, \ldots, 0, 1\} \rightarrow$ spinor $\Delta^{(2)}$

$B_l \equiv$SO$(2l+1)$ - Two elementary UIR's:

$\{N_a\} = \{1, 0, \ldots, 0\} \rightarrow$ vector UIR, D;

$\{N_a\} = \{0, 0, \ldots, 0, 1\} \rightarrow$ spinor Δ

$C_l \equiv$USp$(2l)$ - Two elementary UIR's :

$\{N_a\} = \{1, 0, \ldots, 0\} \rightarrow$ defining UIR D;

$\{N_a\} = \{0, \ldots, 0, 1\} \rightarrow l$th rank antisymmetric "traceless" tensors over D.

$A_l \equiv$SU$(l+1)$ - Two elementary UIR's:

$\{N_a\} = \{0, 0, \ldots, 0, 1\} \rightarrow$ defining "vector" UIR D;

$\{N_a\} = \{1, 0, \ldots, 0\} \rightarrow l$th rank antisymmetric tensors over D.

However, in the C_l and A_l cases, it makes good sense to say that there is only one elementary UIR, and that is the defining UIR D!

16.7 Structure of States within a UIR

We have seen how the UIR's of a compact simple Lie algebra can as a family be classified, and also how they can be built up starting from either fundamental or elementary UIR's. We worked out all the relevant details for the classical nonexceptional families of groups. Now let us briefly look "inside" a generic UIR $\mathcal{D}$, to get a feeling for the kinds of problems and structures involved.

Given a CSLA $\mathcal{L}$ of rank l and order r; what is the dimension of the UIR $\{N_a\}$? Suffice it to say that there do exist explicit formulas due to Weyl, both for the nonexceptional and the exceptional cases. The former expressions are given in Salam's lectures, the latter are included in the book by Wybourne, besides many other places.

Let $\mathcal{W}$ be the set of all weights occurring in $\mathcal{D}$, $\underline{\Lambda}$ the highest weight, and $\underline{\mu}$ a general one. While $\underline{\Lambda}$ is simple, in general $\underline{\mu}$ is not: this is the multiplicity problem. In the most general UIR $\{N_a\}$, it turns out that a basis vector in $\mathcal{V}$ needs $\frac{1}{2}(r - 3l)$ additional independent state labels or quantum numbers to supplement the l weight vector components μ_a which are the eigenvalues of the diagonal generators $H_a, a = 1, 2, \ldots, l$. Thus, in general, a complete commuting set within a UIR consists of $\frac{1}{2}(r - l)$ operators, l of them being generators, and the remainder (in principle) functions of generators (but of course *not* Casimir operators). In some particular UIR's, it could happen that

there is no multiplicity problem – for instance this was the case in all the defining UIR's.

For $A_l = \mathrm{SU}(l+1)$ and for $\mathrm{SO}(n)$ comprising both $B_l = \mathrm{SO}(2l+1)$ and $D_l = \mathrm{SO}(2l)$, there are natural or "canonical" ways of choosing these $\frac{1}{2}(r-3l)$ additional diagonal operators: they are based on the fact that if a UIR of $\mathrm{SU}(l+1)$ (respectively $\mathrm{SO}(n)$) is reduced with respect to the subgroup $\mathrm{SU}(l)\times \mathrm{U}(l)$ [respectively $\mathrm{SO}(n-1)$], each UIR of the latter which occurs does so just once. So one can use the Casimir operators (i.e., basically the UIR labels $\{N\}$) of the chain of sub groups $\mathrm{SU}(l+1) \supset \mathrm{SU}(l) \supset \mathrm{SU}(l-1)\ldots \supset \mathrm{SU}(2)$ in the unitary case, and of $\mathrm{SO}(n) \supset \mathrm{SO}(n-1) \supset \mathrm{SO}(n-2)\ldots \supset \mathrm{SO}(3)$ in the orthogonal case, to solve the multiplicity and state-labelling problem.

Apart from this, what can one say in general terms about the properties of a weight $\underline{\mu} \in \mathcal{W}$? Clearly, by the application of the $\mathrm{SU}(2)^{(\alpha)}$ and $\mathrm{SU}(2)^{(a)}$ subalgebras,

$$\underline{\mu} \in \mathcal{W} \Rightarrow 2\underline{\mu}\cdot\underline{\alpha}^{(a)}/|\underline{\alpha}^{(a)}|^2 \equiv \underline{\mu}\cdot\hat{\alpha}^{(a)} = n_a = 0, \pm 1, \pm 2, \ldots \tag{16.69}$$

In other words, all the states $|\underline{\mu}\ldots\rangle$ for a given $\underline{\mu} \in \mathcal{W}$ are simultaneous eigenstates of the l operators $J_3^{(\alpha^{(a)})}$. Given the integers $\{n_a\}, \underline{\mu}$ is uniquely determined just as $\underline{\Lambda}$ was by $\{N_a\}$:

$$\underline{\mu}\cdot\hat{\alpha}^{(a)} = n_a \Leftrightarrow \underline{\mu} = \sum_{a=1}^{l} n_a \underline{\Lambda}^{(a)} \tag{16.70}$$

But for a given UIR $\{N_a\}$, what $\underline{\mu}$ and so what $\{n\}$ arise?

The general answer is as follows. Every $\underline{\mu} \in \mathcal{W}$ is obtained from $\underline{\Lambda}$ by subtracting a unique non-negative linear combination of the simple roots $\underline{\alpha}^{(a)}$ with integer coefficients:

$$\begin{aligned} &\underline{\Lambda} = \text{highest weight}, \underline{\mu} \in \mathcal{W} \Rightarrow \\ &\underline{\mu} = \underline{\Lambda} - \sum_{a=1}^{l} \nu_a \underline{\alpha}^{(a)}, \quad \nu_a \text{ unique, integer}, \geq 0. \end{aligned} \tag{16.71}$$

The uniqueness follows from the linear independence of the $\underline{\alpha}^{(a)}$. These ν_a are related to the n_a (find the relation!). Now knowing $\underline{\Lambda}$, how do we find out what can be subtracted to get allowed weights present in $\mathcal{D}$? There is a recursive process: we express $\mathcal{W}$ as the union of subsets,

$$\mathcal{W} = \mathcal{W}^{(0)} \cup \mathcal{W}^{(1)} \cup \mathcal{W}^{(2)} \cup \ldots \tag{16.72}$$

Here,

$$\begin{aligned}
&\mathcal{W}^{(0)} = \underline{\Lambda} = \text{simple highest weight;}\\
&\mathcal{W}^{(1)} = \{\underline{\mu} \in \mathcal{W} | \sum_{a=1}^{l} \nu_a = 1\};\\
&\mathcal{W}^{(2)} = \{\underline{\mu} \in \mathcal{W} | \sum_{a=1}^{l} \nu_a = 2\};\\
&\quad \ldots\ldots\ldots\ldots
\end{aligned} \tag{16.73}$$

So $\mathcal{W}^{(k)}$ is called the kth layer of weights – every $\underline{\mu} \in \mathcal{W}^{(k)}$ is k simple roots away from $\underline{\Lambda}$!

How do we find $\mathcal{W}^{(1)}$? In other words: which $\underline{\alpha}^{(a)}$ can be subtracted from $\underline{\Lambda}$? Evidently, with respect to each $SU(2)^{(a)}$ we see from Eq.(16.19) that $\underline{\Lambda}$ is the "maximum m" state:

$$\underline{\Lambda} \text{ under } SU(2)^{(a)}: \; m = j_a = (1/2)N_a. \tag{16.74}$$

Therefore $N_a \geq 1$ is the necessary and sufficient condition for $\underline{\Lambda} - \underline{\alpha}^{(a)}$ to be present:

$$N_a \geq 1 \Leftrightarrow \underline{\Lambda} - \underline{\alpha}^{(a)} \in \mathcal{W}^{(1)} \subset \mathcal{W} \tag{16.75}$$

In general, let $\underline{\mu} \in \mathcal{W}^{(k)}$, and $\underline{\alpha}^{(a)} \in \mathcal{S}$. The question is: does the next layer $\mathcal{W}^{(k+1)}$ contain $\underline{\mu} - \underline{\alpha}^{(a)}$? The answer is given by an application of $SU(2)^{(a)}$ to $\underline{\mu}$, as one would of course expect. With respect to $SU(2)^{(a)}$, $\underline{\mu}$ corresponds to a magnetic quantum number,

$$\underline{\mu} \text{ under } SU(2)^{(a)}: m = \underline{\mu} \cdot \underline{\alpha}^{(a)} / |\underline{\alpha}^{(a)}|^2 = \frac{1}{2} n_a \tag{16.76}$$

We now see: if the contents of $\mathcal{W}^{(0)}, \mathcal{W}^{(1)}, \ldots, \mathcal{W}^{(k)}$ are known; if

$$\begin{aligned}
&\underline{\mu} \in \mathcal{W}^{(k)}, \; \underline{\mu} + \underline{\alpha}^{(a)} \in \mathcal{W}^{(k-1)}, \; \underline{\mu} + 2\underline{\alpha}^{(a)} \in \mathcal{W}^{(k-2)}, \; \ldots,\\
&\underline{\mu} + p\underline{\alpha}^{(a)} \in \mathcal{W}^{(k-p)}, \; \underline{\mu} + (p+1)\underline{\alpha}^{(a)} \notin \mathcal{W},
\end{aligned}$$

and if

$$m > -(m+p) \quad \text{or} \quad m > -p/2,$$

i.e.: m is not yet the least possible value of the magnetic quantum number, then

$$\underline{\mu} - \underline{\alpha}^{(a)} \in \mathcal{W}^{(k+1)}$$

More formally one can express the situation thus:

$$\begin{aligned}
&\underline{\mu} \in \mathcal{W}^{(k)}, \; \underline{\mu} + \rho\underline{\alpha}^{(a)} \in \mathcal{W} \quad \text{for} \quad \rho = 1, 2, \ldots, p,\\
&\underline{\mu} + (p+1)\underline{\alpha}^{(a)} \notin \mathcal{W}, \; \underline{\mu} \cdot \hat{\alpha}^{(a)} + p > 0 \Rightarrow\\
&\underline{\mu} - \underline{\alpha}^{(a)} \in \mathcal{W}^{(k+1)}
\end{aligned} \tag{16.77}$$

In principle, then, once $\underline{\Lambda}$ is chosen, and knowing $\mathcal{S}$, all the weights present in $\mathcal{D}$ are known.

The *level* of a weight $\underline{\mu}$ is the same as the "layer number" k above, namely the "number of simple roots" to be subtracted from $\underline{\Lambda}$ to get to it. This is of course not the same as the *multiplicity* of $\underline{\mu}$, though all the weights of a given layer or at a given level do have the same multiplicity. This multiplicity is given by a formula due to Freudenthal (quoted for instance in the book by Wybourne). The "highest layer" is the one with maximum k, and this maximum value is denoted by $T(\underline{\Lambda})$. It is also called the "height" of the UIR. The behaviour of multiplicities shows a pyramidal or, better, "spindle-shaped" structure. Multiplicity at level k is the same as at level $T(\underline{\Lambda}) - k$. The highest layer $\mathcal{W}^{(0)}$ has a simple weight $\underline{\Lambda}$, so multiplicity one; then the multiplicity keeps increasing for a while as we come down to lower layers, reaching a maximum at $k = \frac{1}{2}T(\underline{\Lambda})$. After that it keeps decreasing again, until for the lowest weight $\underline{\nu}$ it becomes unity again. Exercise: Construct the simple roots $\mathcal{S}$ and the weight and multiplicity diagrams for the UIR's (1 0), (0, 1), (1, 1), (2, 0), (0, 2) of SU(3). In the process many of the general results described above become clear.

Exercises for Chapter 16

1. For the orthogonal groups $\mathcal{D}_l = \mathrm{SO}(2l), B_l = \mathrm{SO}(2l+1)$, prove that the generators belong to the second rank antisymmetric tensor representations $\mathcal{D}^{(2)}$, and that this is the adjoint representation.

2. Work out, in the SU(3) case, the simple roots $\mathcal{S}$, and the weight and multiplicity diagrams for the UIR's (1, 0), (0,1), (1, 1), (2, 0) and (0, 2).

3. Is the adjoint representation (1, 1) of SU(3) a faithful representation? If not, why not?

Chapter 17

Spinor Representations for Real Orthogonal Groups

We have seen in the previous Chapter that among the fundamental UIR's for the groups $D_l = \mathrm{SO}(2l)$ and $B_l = \mathrm{SO}(2l+1)$ there are some "unusual" representations which cannot be obtained by any finite algebraic means from the familiar defining vector representations of these groups. These are the spinor representations. For D_l we saw that there are two inequivalent spinor UIR's, which we denoted by $\Delta^{(1)}$ and $\Delta^{(2)}$; their descriptions as fundamental UIR's, and their highest weights, were found to be (see Eqs.(16.43),(16.45)):

$$\begin{aligned}\mathcal{D}^{(l-1)} \equiv \Delta^{(1)} &\equiv \{0,\ldots,0,1,0\} : \underline{\Lambda}^{(l-1)} = \frac{1}{2}(\underline{e}_1 + \underline{e}_2 + \ldots + \underline{e}_{l-1} - \underline{e}_l) \\ &= (1/2, 1/2, \ldots, 1/2, -1/2); \\ \mathcal{D}^{(l)} \equiv \Delta^{(2)} &\equiv \{0,\ldots,0,1\} : \underline{\Lambda}^{(l)} = \frac{1}{2}(\underline{e}_1 + \underline{e}_2 + \ldots + \underline{e}_{l-1} + \underline{e}_l) \\ &= (1/2, 1/2, \ldots, 1/2, 1/2) \end{aligned} \tag{17.1}$$

For B_l, there is only one spinor UIR, which we wrote as Δ (see Eqs.(16.50,16.52)):

$$\begin{aligned}\mathcal{D}^{(l)} \equiv \Delta &\equiv \{0,\ldots,0,1\} : \underline{\Lambda}^{(l)} = \frac{1}{2}(\underline{e}_1 + \underline{e}_2 + \ldots + \underline{e}_l) \\ &= (1/2, 1/2, \ldots, 1/2) \end{aligned} \tag{17.2}$$

As part of the general Cartan theory of UIR's and fundamental UIR's of compact simple Lie groups and algebras, these spinor representations more or less fall out automatically, once the root systems and simple roots for D_l and B_l are understood. At the same level one also understands and accepts that for the unitary and the symplectic groups there is nothing analogous to spinors. But from the point of view of physical applications, spinors are unusually interesting and useful quantities, and it is therefore appropriate that we understand them in some detail from a somewhat different starting point. In doing so, of course, we can be guided by the general theory outlined in the previous Chapters.

In the succeeding sections, we shall deal at first with the spinor UIR's for $D_l = \mathrm{SO}(2l)$, and later for $B_l = \mathrm{SO}(2l+1)$. In each case we also clarify some of the relationships between spinors and antisymmetric tensors which were briefly alluded to in the previous Chapter. The entire treatment is based on the representations of the Dirac algebra, many aspects of which were first worked out by Brauer and Weyl. It is possible, and probably most elegant, to base the entire analysis on an unspecified representation of the Dirac algebra, i.e.,to develop everything in a representation independent way. However, taking inspiration from Feynman, we shall do all our calculations in a specific representation, and believe there is no sense of defeat in doing so!

The $\mathrm{SO}(2l)$ and $\mathrm{SO}(2l+1)$ treatments in this Chapter will form a basis for studying spinors for the pseudo-orthogonal groups $\mathrm{SO}(p,q)$ in the next Chapter.

17.1 The Dirac Algebra in Even Dimensions

The introduction of spinors in the context of the Lorentz group, based on the Dirac anticommutation relations which arise in the relativistic electron equation is quite well-known. We start from a similar algebra for arriving at spinors with respect to $\mathrm{SO}(2l)$.

Consider the problem of finding hermitian matrices γ_A of suitable dimension obeying the algebraic relations

$$\begin{aligned} \{\gamma_A, \gamma_B\} \equiv \gamma_A\gamma_B + \gamma_B\gamma_A = 2\delta_{AB}, \\ A, B = 1, 2, \ldots, 2l \end{aligned} \tag{17.3}$$

The range of indices is appropriate to $\mathrm{SO}(2l)$, and as in the previous discussions of this group, we will use split index notation when necessary. In fact, in such notation, Eq.(17.3) appears as

$$\{\gamma_{ar}, \gamma_{bs}\} = 2\delta_{ab}\delta_{rs}; a, b = 1, 2, \ldots, l; r, s = 1, 2. \tag{17.4}$$

It is an important fact that, upto unitary equivalence, there is only one irreducible hermitian set γ_A obeying these relations, and the dimension is 2^l. The proof is given, for instance, in Brauer and Weyl. We will give an explicit construction, but it is instructive to recognise immediately the "source" of the spinor UIR's of $\mathrm{SO}(2l)$ – it lies in the essential uniqueness of the solution to (17.3). For, if S is any matrix in the defining representation D of $\mathrm{SO}(2l)$, i.e., a real orthogonal unimodular $2l$-dimensional matrix, then

$$\gamma'_A = S_{BA}\gamma_B \tag{17.5}$$

is also a solution to Eq.(17.3), if γ_A is. By the uniqueness statement, there exists then a unitary matrix $U(S)$ – determined up to a phase factor – which connects γ'_A to γ_A, and as is easily seen, which again up to a phase gives a representation of $\mathrm{SO}(2l)$:

$$\begin{aligned} S \in \mathrm{SO}(2l): \; & U(S)\gamma_A U(S)^{-1} = S_{BA}\gamma_B & \text{(a)} \\ & U(S')U(S) = (\text{phase})U(S'S) & \text{(b)} \end{aligned} \tag{17.6}$$

The $U(S)$ contain the spinorial UIR's of SO($2l$), and this is their origin. We will construct their infinitesimal generators in the sequel.

To construct a solution to (17.3) in a convenient form, we take l independent (i.e. mutually commuting) sets of Pauli matrices $\sigma_j^{(a)}, j = 1,2,3, a = 1,2,\ldots,l$. Since each Pauli triplet needs a two dimensional vector space to act upon, the entire set is defined on a vector space $\mathcal{V}$ of dimension 2^l. The construction of γ_A is best expressed in split index notation as

$$\gamma_A \equiv \gamma_{ar} = \sigma_r^{(a)}\, \sigma_3^{(a+1)}\, \sigma_3^{(a+2)} \ldots \sigma_3^{(l)}, \qquad r = 1,2, a = 1,2,\ldots,l \tag{17.7}$$

Notice that γ_{ar} contains no contributions (factors) from $\underline{\sigma}^{(1)}, \underline{\sigma}^{(2)}, \ldots, \underline{\sigma}^{(a-1)}$ (However the 2×2 unit matrices in those slots are not indicated but left implicit).

The hermiticity of γ_A is obvious from that of the σ's. The irreducibility is also quite easy to prove: we will be showing in a later section that each $\sigma_3^{(a)}$ can be expressed as a product of two of the γ's (in fact this is already evident from Eq.(17.7)); from this fact we then see immediately that $\sigma_1^{(a)}$ and $\sigma_2^{(a)}$ can also be expressed as a product of γ's. Since this is so, and since the complete set of all l Pauli triplets is certainly irreducible, the irreducibility of the γ_A follows. Checking that the Dirac algebra is satisfied is easy; it is best to look at $a < b$ and $a = b$ separately:

$$\begin{aligned}
a < b:\ & \gamma_{ar}\gamma_{bs} = \sigma_r^{(a)}\, \sigma_3^{(a+1)} \ldots \sigma_3^{(b-1)} \sigma_3^{(b)} \sigma_s^{(b)}, \\
& \gamma_{bs}\gamma_{ar} = \sigma_r^{(a)}\, \sigma_3^{(a+1)} \ldots \sigma_3^{(b-1)} \sigma_s^{(b)} \sigma_3^{(b)} \Rightarrow \\
& \gamma_{ar}\gamma_{bs} = -\gamma_{bs}\gamma_{ar}; \\
a = b:\ & \gamma_{ar}\gamma_{as} = \sigma_r^{(a)}\sigma_s^{(a)} \Rightarrow \\
& \{\gamma_{ar}, \gamma_{as}\} = 2\delta_{rs}, \text{ no sum on } a
\end{aligned} \tag{17.8}$$

For uniqueness up to unitary equivalence, the reader may refer to Brauer and Weyl.

All our further calculations will be based on the above specific 2^l dimensional solution for γ_A acting on $\mathcal{V}$ of dimension 2^l. In some contexts it is necessary to pass from the Dirac algebra for one value of l to that for the next value $l+1$. For this purpose we may note that

$$\begin{aligned}
\gamma_A^{(l+1)} &= \gamma_A^{(l)} \sigma_A^{(l+1)}, \ A = 1,2,\ldots,2l; \\
\gamma_{2l+1 \text{ or } 2l+2}^{(l+1)} &= \sigma_{1 \text{ or } 2}^{(l+1)}
\end{aligned} \tag{17.9}$$

Here for clarity we have indicated as a superscript on γ the relevant "l-value".

17.2 Generators, Weights and Reducibility of $U(S)$ – the spinor UIR's of D_l

The unitary operators $U(S)$ acting on $\mathcal{V}$ are not completely defined by Eq.(17.6). To the same extent, the corresponding generators M_{AB} are also known only upto additive constants. It turns out that both $U(S)$ and M_{AB} can be so chosen as to achieve at most a sign ambiguity in the representation condition in Eq.(17.6) – so we have here a "two-valued" representation of SO($2l$).

The generators of $U(S)$ are

$$M_{AB} = \frac{i}{4}[\gamma_A, \gamma_B] \tag{17.10}$$

These are hermitian, reducible (as we shall soon see), and obey the commutation relations (14.6) as a consequence of the Dirac anticommutation relations (17.3). In checking this, one finds that the numerical factor $\frac{i}{4}$ is essential. In addition to the D_l commutation relations, between M_{AB} and γ_C one has the useful relation

$$[M_{AB}, \gamma_C] = i(\delta_{BC}\gamma_A - \delta_{AC}\gamma_B) \tag{17.11}$$

This is in fact the infinitesimal form of Eq.(17.6a).

For later purposes we will need the analogue to Eq.(17.9), a relation between the M_{AB}'s generating SO($2l$), and the next higher set of M_{AB}'s generating SO($2(l+1)$). If we once again indicate the l-value as a superscript, we have, using Eq.(17.9) and leaving implicit the unit matrices in relevant subspaces:

$$\begin{aligned} A, B &= 1, 2, \ldots, 2l \\ M_{AB}^{(l+1)} &= M_{AB}^{(l)} \\ M_{A,2l+1}^{(l+1)} &= -\frac{1}{2}\gamma_A^{(l)}\sigma_2^{(l+1)}; \\ M_{A,2l+2}^{(l+1)} &= +\frac{1}{2}\gamma_A^{(l)}\sigma_1^{(l+1)} \\ M_{2l+1,2l+2}^{(l+1)} &= -\frac{1}{2}\sigma_3^{(l+1)} \end{aligned} \tag{17.12}$$

Going back to the M_{AB} for a given fixed l, we can identify the Cartan subalgebra generators H_a:

$$H_a = M_{a1,a2} = \frac{i}{2}\gamma_{a1}\gamma_{a2} = -\frac{1}{2}\sigma_3^{(a)}, a = 1, 2, \ldots, l \tag{17.13}$$

Therefore, in the usual representation of Pauli matrices, the H_a are already simultaneously diagonal, and we immediately have an orthonormal basis for $\mathcal{V}$ given by vectors with definite nondegenerate weights. We use ϵ_a for the two possible eigenvalues of each H_a, so we describe this basis for $\mathcal{V}$ in this way:

$$\begin{aligned} &|\{\epsilon\}\rangle = |\{\epsilon_1, \epsilon_2, \ldots, \epsilon_l\}\rangle, \epsilon_a = \pm 1; \\ &H_a|\{\epsilon\}\rangle = \frac{1}{2}\epsilon_a|\{\epsilon\}\rangle \\ &\sigma_3^{(a)}|\{\epsilon\}\rangle = -\epsilon_a|\{\epsilon\}\rangle \end{aligned} \tag{17.14}$$

The weight vector $\underline{\mu}$ associated with $|\{\epsilon\}\rangle$ is evidently

$$\underline{\mu} = \frac{1}{2}(\epsilon_1 \underline{e}_1 + \epsilon_2 \underline{e}_2 + \ldots + \epsilon_l \underline{e}_l) \tag{17.15}$$

So the complete set of weights occurring in $\mathcal{V}$ is

$$\frac{1}{2}(\pm \underline{e}_1 \pm \underline{e}_2 \pm \cdots \pm \underline{e}_l)$$

and each being simple, the total number 2^l is precisely the dimension of $\mathcal{V}$.

Now we come to the question of the reducibility of the unitary representation $U(S)$ of D_l. Symbolically the relation between these finite unitary operators and the generators is

$$\begin{aligned} U(S) &\sim \exp\left(-\frac{i}{2}\omega_{AB} M_{AB}\right) \\ \omega_{AB} &= -\omega_{BA} = \text{real parameters.} \end{aligned} \tag{17.16}$$

The operators $U(S)$ happen to be reducible because there exists a nontrivial matrix γ_F which, by virtue of anticommuting with each γ_A, commutes with the generators M_{AB}. All this happens because we are dealing with an even number of dimensions, $2l$. (In analogy with the properties of the matrix γ_5 in the Dirac electron theory, we write F as a subscript on γ_F, to remind us of *FIVE*). We define

$$\begin{aligned} \gamma_F &= i^l \gamma_1 \gamma_2 \ldots \gamma_{2l} \\ &= (-1)^l \sigma_3^{(1)} \sigma_3^{(2)} \ldots \sigma_3^{(l)}, \end{aligned} \tag{17.17}$$

and immediately verify all the following basic properties:

$$\begin{aligned} &\gamma_F^\dagger = \gamma_F^T = \gamma_F^* = \gamma_F^{-1} = \gamma_F && \text{(a)} \\ &\{\gamma_A, \gamma_F\} = 0, \\ &[M_{AB}, \gamma_F] = 0, \\ &U(S)\gamma_F = \gamma_F U(S); && \text{(b)} \\ &\gamma_F |\{\epsilon\}\rangle = \left(\prod_{a=1}^{l} \epsilon_a\right) |\{\epsilon\}\rangle && \text{(c)} \end{aligned} \tag{17.18}$$

Since γ_F is already diagonal, we call the construction (17.7) a Weyl representation of the γ's.

We can split $\mathcal{V}$ into the direct sum of two 2^{l-1} dimensional subspaces $\mathcal{V}_1, \mathcal{V}_2$ corresponding to $\gamma_F = \pm 1$:

$$\begin{aligned} &\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2, \\ &\gamma_F = -1 \text{ on } \mathcal{V}_1, +1 \text{ on } \mathcal{V}_2 \end{aligned} \tag{17.19}$$

Clearly, a basis for $\mathcal{V}_1$ consists of all $|\{\epsilon\}\rangle$ with an odd number of ϵ's taking the value -1, while for $\mathcal{V}_2$ we use $|\{\epsilon\}\rangle$ with an even number of ϵ's being -1. The unitary representation $U(S)$ of D_l restricted to $\mathcal{V}_1$ then gives us the UIR $\Delta^{(1)}$ with the highest weight being

$$\underline{\Lambda}^{(l-1)} = \frac{1}{2}(\underline{e}_1 + \underline{e}_2 + \cdots + \underline{e}_{l-1} - \underline{e}_l) \tag{17.20}$$

We see here the effect of having to include at least one minus sign among the ϵ's. A spinor (an element of $\mathcal{V}_1$) transforming according to this spinor UIR of $SO(2l)$ is called a *left handed* spinor. On the other hand, on $\mathcal{V}_2$ the $U(S)$ furnish us with the other spinor UIR $\Delta^{(2)}$ characterised by the highest weight

$$\underline{\Lambda}^{(l)} = \frac{1}{2}(\underline{e}_1 + \underline{e}_2 + \cdots + \underline{e}_l) \tag{17.21}$$

Elements of $\mathcal{V}_2$ are then called *right handed* spinors. We can express this reduction of the representation $U(S)$ of D_l symbolically as:

$$\begin{aligned} U(\cdot) &= \text{reducible spinor representation of } D_l \text{ on } \mathcal{V} \text{ generated by } M_{AB} \\ &= \text{lefthanded spinor UIR } \Delta^{(1)} \text{on } \mathcal{V}_1 \\ &\quad \oplus \text{righthanded spinor UIR } \Delta^{(2)} \text{ on } \mathcal{V}_2 \end{aligned} \tag{17.22}$$

While the M_{AB} and $U(S)$ are thus reducible the γ_A are of course irreducible on $\mathcal{V}$: in fact since they each anticommute with γ_F, the only non-zero matrix elements of γ_A are those connecting $\mathcal{V}_1$ to $\mathcal{V}_2$ and vice versa!

17.3 Conjugation Properties of Spinor UIR's of D_l

Each of the representations $U(\cdot)$, $\Delta^{(1)}$ and $\Delta^{(2)}$ is unitary, i.e., self adjoint. We would like to know how they each behave under complex conjugation. Here the pattern keeps changing systematically from one l-value to the next, which is why we assembled the Eqs.(17.9),(17.12).

We ask if the reducible representation $U(\cdot)$ is equivalent to its complex conjugate, i.e., whether there is a matrix C acting on $\mathcal{V}$ such that

$$\begin{aligned} U(S)^* &= CU(S)C^{-1}, \\ i.e., \quad (U(S)^T)^{-1} &= CU(S)C^{-1}, \\ i.e., \quad -M_{AB}^T &= CM_{AB}C^{-1} \end{aligned} \tag{17.23}$$

It turns out that such a matrix C does exist, and it can be constructed recursively from one value of l to the next. Of course since $U(\cdot)$ is reducible C will not be unique. We will construct one possible C matrix in a convenient form, see how its properties change with l, and also examine its properties with

respect to γ_F. This last will then disclose the conjugation behaviours of the UIR's $\Delta^{(1)}$ and $\Delta^{(2)}$.

Since the building up of C varies from l to $l+1$, momentarily let us reinstate the l-value as a superscript on all relevant matrices and factors. The desired behaviour of $M_{AB}^{(l)}$ under conjugation by $C^{(l)}$ suggests that we ask for the $\gamma_A^{(l)}$ to transform as follows:

$$C^{(l)}\gamma_A^{(l)}(C^{(l)})^{-1} = \epsilon_l(\gamma_A^{(l)})^T, \epsilon_l = \pm 1 \tag{17.24}$$

For either value of ϵ_l, we do get $M_{AB}^{(l)}$ transforming correctly. Suppose $C^{(l)}$ has been found; let us see if $C^{(l+1)}$ can be taken as $C^{(l)}$ times a suitable 2×2 matrix in the space of the additional Pauli triplet $\underline{\sigma}^{(l+1)}$:

$$\begin{aligned} C^{(l+1)} &= C^{(l)}A^{(l+1)}, \\ A^{(l+1)} &= \text{function of } \underline{\sigma}^{(l+1)} \end{aligned} \tag{17.25}$$

The requirement

$$C^{(l+1)}M_{AB}^{(l+1)}(C^{(l+1)})^{-1} = -(M_{AB}^{(l+1)})^T,\ A, B = 1, 2, \ldots, 2l+2, \tag{17.26}$$

given Eqs.(17.9),(17.12),(17.24),(17.25), leads to conditions on $A^{(l+1)}$ as

$$\begin{aligned} &\text{(a) } A, B \leq 2l: \text{ no conditions;} \\ &\text{(b) } A \leq 2l, B = 2l+1: \ A^{(l+1)}\sigma_2^{(l+1)} = \epsilon_l\sigma_2^{(l+1)}A^{(l+1)}; \\ &\text{(c) } A \leq 2l, B = 2l+2: \ A^{(l+1)}\sigma_1^{(l+1)} = -\epsilon_l\sigma_1^{(l+1)}A^{(l+1)}; \\ &\text{(d) } A = 2l+1, B = 2l+2: \ A^{(l+1)}\sigma_3^{(l+1)} = -\sigma_3^{(l+1)}A^{(l+1)}; \end{aligned} \tag{17.27}$$

These conditions fix $A^{(l+1)}$ depending on ϵ_l:

$$\begin{aligned} \epsilon_l = +1: \ A^{(l+1)} &= i\sigma_2^{(l+1)} \\ \epsilon_l = -1: \ A^{(l+1)} &= \sigma_1^{(l+1)} \end{aligned} \tag{17.28}$$

(we have opted to make $A^{(l+1)}$ real). But then what about ϵ_{l+1} in

$$C^{(l+1)}\gamma_A^{(l+1)}(C^{(l+1)})^{-1} = \epsilon_{l+1}(\gamma_A^{(l+1)})^T,\ A = 1, 2, \ldots, 2l+2? \tag{17.29}$$

Now we discover:

$$\begin{aligned} &\text{(a) } A \leq 2l: \ \epsilon_{l+1} = -\epsilon_l \\ &\text{(b) } A = 2l+1: \ A^{(l+1)}\sigma_1^{(l+1)} = \epsilon_{l+1}\sigma_1^{(l+1)}A^{(l+1)} \\ &\text{(c) } A = 2l+2: \ A^{(l+1)}\sigma_2^{(l+1)} = -\epsilon_{l+1}\sigma_2^{(l+1)}A^{(l+1)} \end{aligned} \tag{17.30}$$

These requirements are consistent with the previous ones, so this recursive method of constructing $C^{(l)}$ has succeeded. Collecting the results, we have

$$\begin{aligned} C^{(l)} &= A^{(1)}A^{(2)}\ldots A^{(l)} \\ A^{(1)} &= i\sigma_2^{(1)}, A^{(2)} = i\sigma_1^{(2)}, A^{(3)} = i\sigma_2^{(3)}, \ldots \\ \epsilon_l &= (-1)^l \end{aligned} \tag{17.31}$$

We repeat that Eq.(17.23) alone would leave $C^{(l)}$ non-unique since the representation $U(\cdot)$ is reducible. The additional natural requirement (17.24), and the recursive method of construction has led to a possible and convenient choice.

It is now an easy matter to check the following additional properties of $C^{(l)}$:

$$\begin{aligned} &(C^{(l)})^T = (C^{(l)})^{-1} = (-1)^{l(l+1)/2}C^{(l)}, \\ &(C^{(l)})^* = C^{(l)}, \\ &C^{(l)}\gamma_F = (-1)^l\gamma_F C^{(l)} \end{aligned} \tag{17.32}$$

The behaviours of $\Delta^{(1)}$ and $\Delta^{(2)}$ under conjugation can now be read off. For even l, each of $\Delta^{(1)}$ and $\Delta^{(2)}$ is self conjugate; whether they are potentially real or only pseudo-real then depends on the symmetry or antisymmetry of C. For odd $l, \Delta^{(1)}$ and $\Delta^{(2)}$ are mutually conjugate. We can thus draw up the following conjugation table for the spinor UIR's of D_l:

l	$(C^{(l)})^T$	$C^{(l)}$ *vs* γ_F	$\Delta^{(1)}$	$\Delta^{(2)}$
$4m$	$C^{(l)}$	Commute	Potentially real	Potentially real
$4m+1$	$-C^{(l)}$	Anticommute	$\sim \Delta^{(2)*}$	$\sim \Delta^{(1)*}$
$4m+2$	$-C^{(l)}$	Commute	Pseudo-real	Pseudo-real
$4m+3$	$C^{(l)}$	Anticommute	$\sim \Delta^{(2)*}$	$\sim \Delta^{(1)*}$

17.4 Remarks on Antisymmetric Tensors Under $D_l = \mathrm{SO}(2l)$

In Section 16.5 we saw that all but two of the antisymmetric tensor representations of $D_l = \mathrm{SO}(2l)$ occur in the list of fundamental UIR's. However the antisymmetric tensors of ranks $l-1$ and l do not appear in such a role. We devote this section to a brief discussion of such tensors, and a clarification of the products of spinor UIR's in which they occur.

An antisymmetric tensor of rank m, say, is an m-index object $T_{A_1A_2\ldots A_m}$ completely antisymmetric in the indices, which transforms under $S \in \mathrm{SO}(2l)$ in

the following manner:

$$T'_{A_1 A_2 \ldots A_m} = S_{A_1 B_1} S_{A_2 B_2} \ldots S_{A_m B_m} T_{B_1 B_2 \ldots B_m} \tag{17.33}$$

Since for $\mathrm{SO}(2l)$ we have an invariant Levi-Civita tensor $\epsilon_{A_1 A_2 \ldots A_{2l}}$ (normalised to $\epsilon_{1\ 2\ \ldots 2l} = +1$), we need to consider only antisymmetric tensors of ranks $1, 2, \ldots, l-1, l$. Of these, those of ranks $1, 2, \ldots, l-2$ furnish the first $(l-2)$ fundamental UIR's of the group.

We have already seen in Section 16.5 that the highest weight in the representation given by antisymmetric tensors of rank $(l-1)$ is

$$\begin{aligned} \underline{\Lambda} &= \underline{e}_1 + \underline{e}_2 + \cdots + \underline{e}_{l-1} \\ &= \underline{\Lambda}^{(l-1)} + \underline{\Lambda}^{(l)} \end{aligned} \tag{17.34}$$

Thus, this UIR occurs in the reduction of the direct product $\Delta^{(1)} \times \Delta^{(2)}$. Now let us examine antisymmetric tensors of rank l in some detail. Given such a tensor T, we can view the Levi-Civita symbol as defining a linear operator $\hat{\epsilon}$ taking T to T' in the following way:

$$\begin{aligned} T' &= \hat{\epsilon} T : \\ T'_{A_1 \cdots A_l} &= \frac{1}{l!} \epsilon_{A'_1 \cdots A'_l A_1 \cdots A_l} T_{A'_1 \cdots A'_l} \end{aligned} \tag{17.35}$$

It is easy to check that

$$\hat{\epsilon}^2 = (-1)^l \tag{17.36}$$

This permits the definition of self-dual and antiself dual antisymmetric tensors of rank l, as eigentensors of $\hat{\epsilon}$; while this can be done in the real domain if l is even, one has to go to the complex domain if l is odd. For uniform appearance of later results we choose the definitions,

$$\begin{aligned} &\text{Self dual tensors: } \hat{\epsilon} T = i^{l^2} T = (1 \text{ or } i)\, T, \\ &\qquad\qquad \text{according as } l \text{ is even or odd} \qquad \text{(a)} \\ &\text{Antiself dual tensors: } \hat{\epsilon} T = -i^{l^2} T = (-1 \text{ or } -i)\, T, \\ &\qquad\qquad \text{according as } l \text{ is even or odd} \qquad \text{(b)} \end{aligned} \tag{17.37}$$

The existence of the operator $\hat{\epsilon}$ shows that the representation of $\mathrm{SO}(2l)$ given by rank l antisymmetric tensors is reducible: it splits into two UIR's given respectively by self dual and antiself dual tensors. We wish to find the weights that occur in each of these UIR's and in particular their highest weights.

For further algebraic work it is convenient to introduce a formal set of ket vectors on which $\hat{\epsilon}$ and M_{AB} can act, rather than continue to deal with tensors and their numerical components. Thus we shall have kets $|A_1 A_2 \cdots A_l\rangle$, with

each A in the range $1, 2, \ldots, 2l$, subject to the following laws:

$$
\begin{aligned}
&\text{(i) For any permutation } P \in S_l,\\
&\quad |P(A_1 A_2 \ldots A_l)\rangle = \epsilon_P |A_1 A_2 \ldots A_l\rangle,\\
&\quad \epsilon_P = \text{parity of } P = \pm 1\\
&\text{(ii) } M_{AB}|A_1 A_2 \ldots A_l\rangle = i(\delta_{BA_1}|A_1 A_2 \ldots A_l\rangle - \delta_{AA_1}|BA_2 \ldots A_l\rangle\\
&\quad + \delta_{BA_2}|A_1 A \ldots A_l\rangle - \delta_{AA_2}|A_1 B \ldots A_l\rangle \ldots)\\
&\text{(iii) } \hat{\epsilon}|A_1 A_2 \ldots A_l\rangle \equiv \frac{1}{l!}\epsilon_{A_1 \cdots A_l A'_1 \cdots A'_l}|A'_1 A'_2 \ldots A'_l\rangle\\
&\quad = \eta |A_1 A_2 \ldots A_l\rangle,\\
&\quad \eta = i^{l^2} = 1 \text{ or } i \text{ acc. as } l \text{ even or odd for self dual case},\\
&\quad \eta = -i^{l^2} = -1 \text{ or } -i \text{ acc. as } l \text{ even or odd for antiself dual case}
\end{aligned}
\tag{17.38}
$$

While one can certainly introduce inner products among these kets, we do not need to do so. The first property above – a "Pauli Principle" – tells us that in any string $A_1 A_2 \ldots A_l$, we can always assume that all entries are distinct; thus each relevant string is a subset of l distinct integers drawn from $1, 2, \ldots, 2l$. The signs chosen in the action of M_{AB} are consistent with the SO($2l$) commutation relations (14.6). In the space of the above kets, we want to find simultaneous eigenkets of $H_1, H_2, \ldots, H_l$ and see what the corresponding weights $\underline{\mu} = \{\mu_a\}$ look like.

For brevity, let us denote by $\underline{A}$ the string of indices $A_1 A_2 \ldots A_l$ occurring inside each basic ket. In a given $\underline{A}$, a certain pair $2a-1, 2a$ out of $1, 2, \ldots, 2l$ may not occur at all; or one member of this pair may be present but not the other; or finally both members of the pair may be present. Use of Eq.(17.38)(ii) shows that in the first and last cases, we obtain the eigenvalue zero for H_a:

$$
\begin{aligned}
&2a-1, 2a \notin \underline{A} : H_a|\underline{A}\rangle = 0;\\
&2a-1, 2a \in \underline{A} : H_a|\underline{A}\rangle = 0.
\end{aligned}
\tag{17.39}
$$

In fact, the only possible eigenvalues of each H_a (in the space of tensors we are considering here) are $0, \pm 1$. Examples of eigenvectors of, say, $H_1 = M_{12}$ with eigenvalues ± 1 are easily built:

$$
\begin{aligned}
|1A_2 \ldots A_l\rangle \mp i|2A_2 \ldots A_l\rangle \equiv &|1 \mp i2, A_2 \ldots A_l\rangle,\\
&1, 2 \notin A_2 \ldots A_l,\\
H_1|1 \mp i2, A_2 \ldots A_l\rangle = &\pm |1 \mp i2, A_2 \ldots A_l\rangle
\end{aligned}
\tag{17.40}
$$

Here we have introduced a compact way of expressing certain linear combinations of the basic kets $|\underline{A}\rangle$, that will be convenient in what follows. In any weight $\underline{\mu}$ which occurs, then, each $\mu_a = \pm 1$ or 0.

We must now see in a general way how to connect up a vector $|\underline{\mu}\ldots\rangle$, a simultaneous eigenket of the H_a, with the $|\underline{A}\rangle$. For a while we will work only with Eq.(17.38)(i), (ii), and only later impose the $\hat{\epsilon}$-related condition. In a basis ket $|\underline{A}\rangle$, suppose the string $\underline{A}$ is characterised by l_0 completely absent pairs, l_1 pairs contributing only one member each, and l_2 pairs fully present. We easily see that

$$\begin{aligned} l_0 + l_1 + l_2 &= \text{total number of pairs} = l, \\ l_1 + 2l_2 &= \text{number of entries in } \underline{A} = l, \\ \text{i.e.} \qquad l_0 &= l_2 \end{aligned} \tag{17.41}$$

On the other hand, suppose in a particular weight vector $\underline{\mu} = (\mu_1 \ldots \mu_l)$, the entries $0, 1, -1$ occur n_0, n_1 and n_{-1} times respectively. Of course $\underline{\mu}$ may be degenerate. In the expression of an eigenket $|\underline{\mu}\ldots\rangle$ as a linear combination of the $|\underline{A}\rangle$, it must be clear because of Eqs.(17.39),(17.40) that the following relations between $n_0, n_{\pm 1}$ on the one hand, and l_0, l_1, l_2 on the other must be maintained in each term:

$$\begin{aligned} n_0 = l_0 + l_2 = 2l_0 &= \text{even}, \\ n_1 + n_{-1} &= l_1 \end{aligned} \tag{17.42}$$

One more restriction on allowed weights $\underline{\mu}$ emerges: the entry zero must occur an *even* number of times! One can now see why in general a weight $\underline{\mu}$ has multiplicity greater than one, and also when $\underline{\mu}$ will be simple. For each H_a for which $\mu_a = \pm 1$, the disposition of the indices $2a-1, 2a$ in the corresponding pair is completely fixed by Eq.(17.40), and there is no freedom left. But for the even number of H_a's for which $\mu_a = 0$, for each such H_a we have the possibility of either completely dropping or fully including its pair $2a-1, 2a$. In fact, for half of these H_a's we must pick the former alternative and for the other half the latter alternative, since from the connecting relations (17.42) we have $l_0 = l_2 = \frac{1}{2}n_0$! Thinking this through, one realises that $\underline{\mu}$ can have multiplicity only if $n_0 > 0$; and conversely if $n_0 = 0$, i.e., each $\mu_a = \pm 1$, then $\underline{\mu}$ is simple.

As an illustrative example, consider the weight vector $\underline{\mu}$ to be $(1, 1, \ldots, 1, -1, -1, \ldots, -1, 0, \ldots)$:

$$\begin{aligned} \mu_1 = \mu_2 = \cdots = \mu_{n_1} &= 1, \\ \mu_{n_1+1} = \mu_{n_1+2} = \cdots = \mu_{n_1+n_{-1}} &= -1, \\ \mu_a = 0 \text{ for } a &= n_1 + n_{-1} + 1, \ldots, l \end{aligned} \tag{17.43}$$

Then the following linear combination of the basic kets $|\underline{A}\rangle$ is a possible eigenket $|\underline{\mu}\ldots\rangle$:

$$\begin{aligned} |\underline{\mu}\ldots\rangle \sim |&1 - i2,\ 3 - i4,\ \ldots,\ 2n_1 - 1 - i \cdot 2n_1;\ 2n_1 + 1 + i(2n_1 + 2),\ \ldots, \\ &2n_1 + 2n_{-1} - 1 + i(2n_1 + 2n_{-1}),\ \text{some pairs}\rangle \end{aligned} \tag{17.44}$$

By the words "some pairs" we mean some choice of half of the *last* n_0 pairs in $1, 2, \ldots, 2l-1, 2l$. We see immediately why, if $n_0 = 0$, $\underline{\mu}$ is simple — this last choice does not have to be made, and the ket $|\underline{\mu}\rangle$ is fully determined.

Now let us bring in $\hat{\epsilon}$ and ask for its effect on the ket appearing in Eq.(17.44). Going back to Eq.(17.40) for a moment, we see that

$$\hat{\epsilon}|1 \mp i2, A_2 \ldots A_l\rangle = \frac{\pm i}{(l-1)!}\epsilon_{1A_2\ldots A_l 2A'_2\ldots A'_l}|1 \mp i2, A'_2 \ldots A'_l\rangle \tag{17.45}$$

it being understood that neither 1 nor 2 occur in $A_2 \ldots A_l$. We can extend this to the ket appearing in Eq.(17.44); if for simplicity we omit some of the entries in this ket, we can write:

$$\begin{aligned}\hat{\epsilon}|1 - i2, 3 - i4, \ldots,\ &\text{some pairs}\rangle = i^{n_1}(-i)^{n_{-1}} \\ &\times \epsilon_{1,3,\ldots,2n_1+2n_{-1}-1,2,4,\ldots,2n_1+2n_{-1}},\ \text{pairs} \\ &\times |1 - i2, 3 - i4, \ldots,\ \text{pairs absent on left}\rangle,\ \text{no sums on pairs}\end{aligned} \tag{17.46}$$

It helps here to recognize that each pair of indices $2a-1, 2a$ behaves like a "boson" and can be freely moved past other indices with no minus signs being produced! After bringing the ϵ-component here to normal form, we can rewrite the above equation as

$$\begin{aligned}\hat{\epsilon}|1 - i2; 3 - i4; \ldots;\ &\text{some pairs}\rangle = i^{n_1}(-i)^{n_{-1}}(-1)^{1/2(n_1+n_{-1}-1)(n_1+n_{-1})} \\ &\times |1 - i2, 3 - i4, \ldots,\ \text{pairs absent on left}\rangle\end{aligned} \tag{17.47}$$

We are now essentially done. If every ket obeys Eq.(17.38)(iii)) and so is an eigenket of $\hat{\epsilon}$ with eigenvalue η, then on the possible simple weights $\underline{\mu} = \{\mu_a\}$ in which each $\mu_a = \pm 1$ there is the *restriction*

$$\begin{aligned}&n_0 = 0:\ \underline{\mu}\ \text{simple:} = i^{l-n_{-1}} \cdot (-i)^{n_{-1}} \cdot (-1)^{l(l-1)/2} = \eta \\ &\text{i.e.,} \qquad (-1)^{n_{-1}} = \eta^* \cdot i^{l^2}\end{aligned} \tag{17.48}$$

We will interpret this in a moment, but we also notice from Eq.(17.47) that if a tensor has a definite duality property, then any allowed weight $\underline{\mu}$ with $n_0 = 2$, i.e., *two* of the μ_a being zero, is also a simple weight; nontrivial multiplicities show up only if $n_0 = 4, 6, \ldots$!

Now we use Eq.(17.48) to answer the question: for a definite kind of duality property, what is the allowed highest weight $\underline{\Lambda}$? We see from a combination of Eqs.(17.38),(17.48) that the pattern is as follows:

$$\begin{aligned}&\text{Self dual case: } \eta = i^{l^2}:\ n_{-1} = 0, \underline{\Lambda} = \underline{e}_1 + \underline{e}_2 + \ldots + \underline{e}_{l-1} + \underline{e}_l; \\ &\text{Antiself dual case: } \eta = -i^{l^2}:\ n_{-1} = 1, \underline{\Lambda} = \underline{e}_1 + \underline{e}_2 + \ldots + \underline{e}_{l-1} - \underline{e}_l\end{aligned} \tag{17.49}$$

(We see now the justification for the definitions adopted in Eqs.(17.37)!) We have uniformly the result that the direct product $\Delta^{(1)} \times \Delta^{(1)}$ produces, as the leading piece in its reduction, the UIR of rank l antisymmetric antiself dual tensors; while the direct product $\Delta^{(2)} \times \Delta^{(2)}$ produces the UIR of rank l antisymmetric selfdual tensors. A summary of these and other significant properties of $\mathrm{SO}(2l)$ antisymmetric tensors is given in the table:

Antisymmetric tensors of $\mathrm{SO}(2l)$:

Rank	Highest weight	Status	Remark
1	$\underline{e}_1$	Fundamental UIR $\mathcal{D}^{(1)}$	Defining UIR D
2	$\underline{e}_1 + \underline{e}_2$	Fundamental UIR $\mathcal{D}^{(2)}$	Adjoint UIR
3	$\underline{e}_1 + \underline{e}_2 + \underline{e}_3$	Fundamental UIR $\mathcal{D}^{(3)}$	
...	...	...	
$l-2$	$\underline{e}_1 + \underline{e}_2 + \cdots + \underline{e}_{l-2}$	Fundamental UIR $\mathcal{D}^{(l-2)}$	—
$l-1$	$\underline{e}_1 + \underline{e}_2 + \cdots + \underline{e}_{l-1}$	Leading piece in $\Delta^{(1)} \times \Delta^{(2)}$	UIR
l	Self-dual $\underline{e}_1 + \underline{e}_2 + \cdots + \underline{e}_{l-1} + \underline{e}_l$	Leading piece in $\Delta^{(2)} \times \Delta^{(2)}$	UIR
	Antiself-dual $\underline{e}_1 + \underline{e}_2 + \cdots + \underline{e}_{l-1} - \underline{e}_l$	Leading piece in $\Delta^{(1)} \times \Delta^{(1)}$	UIR

17.5 The Spinor UIR's of $B_l = \mathrm{SO}(2l+1)$

The situation here is somewhat simpler than in the case of D_l. As we have seen earlier, there is just one spinor UIR, with highest weight given in Eq.(17.2). Actually, in comparison with the D_l situation, things are slightly more complicated at the Dirac algebra level, and slightly simpler at the spinor UIR level!

We begin by asking for an irreducible, hermitian solution to the "$(2l+1)$ dimensional" Dirac algebra:

$$\{\gamma_A, \gamma_B\} = 2\delta_{AB}, \quad A, B = 1, 2, \ldots, 2l+1 \tag{17.50}$$

It turns out that now there are two inequivalent solutions, each of dimension 2^l:

(i) We take $\gamma_A, A = 1, 2, \ldots, 2l$ as constructed earlier in the case of D_l, and to them we adjoin

$$\gamma_{2l+1} = \gamma_F \tag{17.51}$$

All of Eq.(17.50) are obeyed, and furthermore the following algebraic relation holds:

$$\gamma_1 \gamma_2 \cdots \gamma_{2l} \gamma_{2l+1} = (-i)^l \tag{17.52}$$

(ii) Take γ_A for $A = 1, 2, \ldots, 2l$ as before, but choose now

$$\gamma'_{2l+1} = -\gamma_F \tag{17.53}$$

The entire set $\gamma'_A \equiv (\gamma_1, \gamma_2, \ldots, \gamma_{2l}, -\gamma_F)$ again is a solution to (17.50); however in place of (17.52) we have a *different* algebraic relation,

$$\gamma'_1\gamma'_2\cdots\gamma'_{2l}\gamma'_{2l+1} = -(-i)^l \tag{17.54}$$

which is why *no* unitary transformation can connect the set γ_A to the set γ'_A. One does however have the connection

$$\gamma'_A = -\gamma_F\gamma_A\gamma_F^{-1}, \quad A = 1, 2, \ldots, 2l+1 \tag{17.55}$$

which we will use shortly. We shall generally work with the representation (i) above.

If $S = (S_{AB}) \in \mathrm{SO}(2l+1)$, by continuity and preservation of the algebra (17.50) we see that $S_{BA}\gamma_B$ must be unitarily related to γ_A, *not to* γ'_A:

$$\begin{aligned} S \in \mathrm{SO}(2l+1) : S_{BA}\gamma_B &= U(S)\gamma_A U(S)^{-1}, \\ U(S')U(S) &= (\text{phase})\, U(S'S). \end{aligned} \tag{17.56}$$

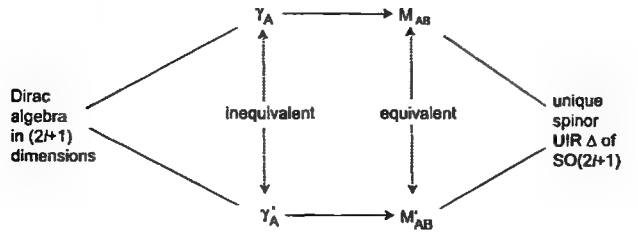
Similarly, we have

$$\begin{aligned} S \in \mathrm{SO}(2l+1) : S_{BA}\gamma'_B &= U'(S)\gamma'_A(U'(S))^{-1}, \\ U'(S')U'(S) &= (\text{phase})\, U'(S'S). \end{aligned} \tag{17.57}$$

The generators M_{AB} for $U(S)$ and M'_{AB} for $U'(S)$ are unitarily related,

$$\begin{aligned} M_{AB} &= \frac{i}{4}[\gamma_A, \gamma_B] \\ M'_{AB} &= \frac{i}{4}[\gamma'_A, \gamma'_B] = \gamma_F M_{AB}\gamma_F^{-1} \\ U'(S) &= \gamma_F U(S)\gamma_F^{-1} \end{aligned} \tag{17.58}$$

It is here that we use Eq.(17.55). Thus, while we do have two inequivalent representations of the Dirac algebra in odd dimensions, they lead to equivalent generators for $\mathrm{SO}(2l+1)$, and so to essentially just one spinor UIR. In fact the irreducibility of $U(S)$ follows from that of M_{AB}, which in turn is a consequence of our being able to express each γ_A as a product of l of the M's. One can depict the situation pictorially thus:



Hereafter let us stick to γ_A, M_{AB} and $U(S)$. The space of this UIR is $\mathcal{V}$ of dimension 2^l, with the basis $|\{\epsilon\}\rangle$ set up in Eq.(17.14). The Cartan subalgebra for $\mathrm{SO}(2l+1)$ is the same as for $\mathrm{SO}(2l)$, and the highest weight is as expected, $\underline{\Lambda} = (1/2, 1/2, \ldots, 1/2)$. As for the behaviour of $U(S)$ under conjugation, it is necessarily self conjugate, and the same C matrix will do as previously; from Eqs.(17.18a), (17.24), (17.31), (17.51) we have uniformly:

$$\begin{aligned} C\gamma_A C^{-1} &= (-1)^l \gamma_A^T, \quad A = 1, 2, \ldots, 2l+1; \\ CM_{AB}C^{-1} &= -M_{AB}^T; \\ (U(S)^T)^{-1} &= CU(S)C^{-1} \end{aligned} \tag{17.59}$$

Since C is sometimes symmetric and sometimes antisymmetric, as determined by Eq.(17.32), $U(S)$ is correspondingly potentially real or pseudoreal:

$$\begin{aligned} l = 4m \text{ or } 4m+3: &\ \Delta \text{ of } \mathrm{SO}(2l+1) \text{ is potentially real;} \\ l = 4m+1 \text{ or } 4m+2: &\ \Delta \text{ of } \mathrm{SO}(2l+1) \text{ is pseudo-real} \end{aligned} \tag{17.60}$$

Combining these results with those of Section 17.3, we see that as a whole the properties of the spinor UIR's of the orthogonal groups exhibit a "cycle of eight" structure.

Group	$\Delta^{(1)}, \Delta^{(2)}$	Δ
$\mathrm{SO}(8m) = D_{4m}$	Real, dim. 2^{4m-1}	—
$\mathrm{SO}(8m+1) = B_{4m}$	—	Real, dim. 2^{4m}
$\mathrm{SO}(8m+2) = D_{4m+1}$	Mutually conjugate, dim. 2^{4m}	—
$\mathrm{SO}(8m+3) = B_{4m+1}$	—	Pseudo-real, dim. 2^{4m+1}
$\mathrm{SO}(8m+4) = D_{4m+2}$	Pseudo-real, dim. 2^{4m+1}	—
$\mathrm{SO}(8m+5) = B_{4m+2}$	—	Pseudo-real, dim. 2^{4m+2}
$\mathrm{SO}(8m+6) = D_{4m+3}$	Mutually conjugate, dim. 2^{4m+2}	—
$\mathrm{SO}(8m+7) = B_{4m+3}$	—	Real, dim. 2^{4m+3}

This table, and the behaviour of dimensionalities, tempts us to conclude this section with the following remarks. Take the spinor UIR Δ of $B_l = \mathrm{SO}(2l+1)$, of dimension 2^l. Each of the sets of matrices γ_A and M_{AB}, for $A, B = 1, 2, \ldots, 2l+$

1, is separately irreducible. Evidently, M_{AB} and γ_A form a Lie algebra; and so do M_{AB} and $-\gamma_A$. These just give the two inequivalent spinor UIR's $\Delta^{(1)}, \Delta^{(2)}$ of $D_{l+1} = \mathrm{SO}(2l+2)$, both of dimension 2^l, in some order. In fact,

$$M_{A,2l+2} = \mp\frac{1}{2}\gamma_A, A = 1, 2, \ldots, 21+1 \Rightarrow$$
$$\underline{\Delta} = \frac{1}{2}(\underline{e}_1 + \underline{e}_2 + \cdots + \underline{e}_l \mp \underline{e}_{l+1}) \Rightarrow$$
$$\Delta^{(1)} \text{ or } \Delta^{(2)} \text{ of } D_{l+1} \tag{17.61}$$

17.6 Antisymmetric Tensors under $B_l = \mathrm{SO}(2l+1)$

Here again the situation is much simpler than with D_l. There is no duality operation on antisymmetric tensors of a given rank. The presence of the Levi-Civita tensor $\epsilon_{A_1,\ldots,A_{2l+1}}$ allows us to limit ourselves to antisymmetric tensors of ranks $1, 2, \ldots, l-1, l$. All but the last are bases for fundamental UIR's. As noted in Section 16.5, the lth rank antisymmetric tensor UIR has as its highest weight twice the highest weight of the spinor UIR Δ. All the pertinent information about B_l antisymmetric tensors can thus be summarised in this fashion:

Antisymmetric tensors of $\mathrm{SO}(2l+1)$:

Rank	Highest weight	Status	Remark
1	$\underline{e}_1$	Fundamental UIR $\mathcal{D}^{(1)}$	Defining UIR D
2	$\underline{e}_1 + \underline{e}_2$	Fundamental UIR $\mathcal{D}^{(2)}$	Adjoint UIR
3	$\underline{e}_1 + \underline{e}_2 + \underline{e}_3$	Fundamental UIR $\mathcal{D}^{(3)}$	—
	.....		
$l-1$	$\underline{e}_1 + \underline{e}_2 + \cdots + \underline{e}_{l-1}$	Fundamental UIR $\mathcal{D}^{(l-1)}$	—
l	$\underline{e}_1 + \underline{e}_2 + \cdots + \underline{e}_l$	Leading piece in $\Delta \times \Delta$	UIR

Chapter 18

Spinor Representations for Real Pseudo Orthogonal Groups

The pseudo orthogonal group $SO(p,q)$ is the connected proper group of real linear transformations in a $(p+q)$-dimensional "space-time" preserving a nondegenerate but indefinite metric with signature $(++\cdots+--\cdots-)$, there being p plus and q minus signs. While its Lie algebra is simple, (provided $(p,q) \neq (2,2)$!), it is a noncompact Lie group, and all its nontrivial unitary representations are infinite dimensional. For many physical applications, however, one is interested in the spinor representations of these groups, and furthermore in various kinds of spinors. These representations are finite dimensional and non-unitary. We can use all the algebraic results we have put together in Chapter 17, to provide a discussion of spinors for $SO(p,q)$. Naturally the behaviours of spinors depend very much on whether the total dimensionality $p+q$ is even or odd, and whether the number of minus signs in the metric is even or odd. We give a concise account of these matters in this Chapter.

18.1 Definition of $SO(q,p)$ and Notational Matters

We shall use Greek indices $\mu, \nu, \ldots$ to go over the full range $1, 2, \ldots, p+q$; early Latin indices $j, k, \ldots$ to go over the positive metric "spatial" dimensions $1, 2, \ldots, p$; and late Latin indices $r, s, \ldots$ to go over the negative metric "time-like" dimensions $p+1, p+2, \ldots, p+q$. The metric tensor $\eta_{\mu\nu}$ is diagonal with

$$\begin{aligned} \eta_{11} = \eta_{22} = \cdots = \eta_{pp} &= +1 \\ \eta_{p+1,p+1} = \cdots = \eta_{p+q,p+q} &= -1 \end{aligned} \tag{18.1}$$

This metric will also be used, along with its inverse $\eta^{\mu\nu}$, for lowering and raising Greek indices.

The group $SO(p,q)$ consists of all real $p+q$ dimensional matrices $\Lambda = (\Lambda^{\mu}{}_{\nu})$ obeying the conditions

$$\begin{aligned} \Lambda^{\mu}{}_{\nu}\Lambda_{\mu\lambda} \equiv \eta_{\mu\rho}\Lambda^{\mu}{}_{\nu}\Lambda^{\rho}{}_{\lambda} = \eta_{\nu\lambda}, \\ \det \Lambda = +1, \end{aligned} \tag{18.2}$$

and in addition the condition that it lie in the component continuously connected to the identity. (Often this last condition is indicated by denoting the group in a more specific way than by just $SO(p,q)$, but we shall leave it implicit). After removal of the conventional "quantum mechanical i", the Lie algebra of $SO(p,q)$ is spanned by $M_{\mu\nu} = -M_{\nu\mu}$ obeying the commutation relations

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(\eta_{\nu\rho}M_{\mu\sigma} - \eta_{\mu\rho}M_{\nu\sigma} + \eta_{\nu\sigma}M_{\rho\mu} - \eta_{\mu\sigma}M_{\rho\nu}) \tag{18.3}$$

In a nontrivial unitary representation of $SO(p,q)$, which is necessarily infinite dimensional, the $M_{\mu\nu}$ would be hermitian, or more precisely self adjoint operators. In finite dimensional non-unitary matrix representations, however, we can assume without loss of generality that

$$\begin{aligned} M^{\dagger}_{jk} &= M_{jk} = \text{hermitian } SO(p) \text{ generators}, \\ M^{\dagger}_{rs} &= M_{rs} = \text{hermitian } SO(q) \text{ generators}, \\ M^{\dagger}_{jr} &= -M_{jr} = \text{antihermitian generators} \end{aligned} \tag{18.4}$$

These properties are of course consistent with Eqs.(18.3).

In working out the spinor representations of D_l and B_l we constructed hermitian irreducible matrices γ_A in Eqs.(17.7), (17.17). We shall now write $\gamma^{(0)}_A$ for them:

$$\begin{aligned} \gamma^{(0)}_A &= \text{previous } \gamma_A, A = 1, 2, \ldots, 2l; \\ \gamma^{(0)}_{2l+1} &= \gamma_F. \end{aligned} \tag{18.5}$$

18.2 Spinor Representations $S(\Lambda)$ of $SO(p,q)$ for $p+q \equiv 2l$

The appropriate Dirac algebra is

$$\{\gamma_\mu, \gamma_\nu\} = 2\eta_{\mu\nu}, \mu, \nu = 1, 2, \ldots, p+q = 2l \tag{18.6}$$

Upto equivalence, there is a unique irreducible representation, on a space $\mathcal{V}$ of dimension 2^l. We take the solution

$$\begin{aligned} \gamma_j &= \gamma^{\dagger}_j = \gamma^{(0)}_j, \quad j = 1, 2, \ldots, p; \\ \gamma_r &= -\gamma^{\dagger}_r = i\gamma^{(0)}_r, r = p+1, \ldots, 2l \end{aligned} \tag{18.7}$$

Thus the hermiticity relations can be expressed as

$$\gamma_\mu^\dagger = \eta_{\mu\mu}\gamma_\mu, \quad \text{no sum} \tag{18.8}$$

With this hermiticity property specified, the solution to (18.6) is unique upto *unitary* equivalence.

For any $\Lambda \in \mathrm{SO}(p,q)$,

$$\gamma'_\mu = \Lambda^\nu{}_\mu \gamma_\nu \tag{18.9}$$

is a solution to (18.6), if γ_μ is. (However in general γ'_μ will *not* enjoy the hermiticity properties (18.8), unless $\Lambda \in \mathrm{SO}(p)\times \mathrm{SO}(q)$). Therefore there must be a similarity transformation connecting γ'_μ and γ_μ as

$$\begin{aligned} \Lambda^\nu{}_\mu \gamma_\nu &= S(\Lambda)\gamma_\mu S(\Lambda)^{-1}, \\ S(\Lambda')S(\Lambda) &= \pm S(\Lambda'\Lambda). \end{aligned} \tag{18.10}$$

This is the origin of the 2^l-dimensional (reducible) "Dirac spinor" representation $S(\Lambda)$ of $\overline{\mathrm{SO}(p,q)}$. It is in fact a double-valued representation, or more precisely, a representation of the universal covering group $\overline{\mathrm{SO}(p,q)}$, which in some cases is more than a double covering of $\mathrm{SO}(p,q)$.

The infinitesimal generators of $S(\Lambda)$ such that

$$\begin{aligned} S(\Lambda) &\sim \exp\left(\frac{i}{2}\omega^{\mu\nu}M_{\mu\nu}\right), \\ \omega_{\mu\nu} &= -\omega_{\nu\mu} = \text{real}, \end{aligned} \tag{18.11}$$

are

$$M_{\mu\nu} = \frac{i}{4}[\gamma_\mu, \gamma_\nu]. \tag{18.12}$$

One easily checks the validity of both Eqs.(18.3), (18.4), and also of

$$[M_{\mu\nu}, \gamma_\rho] = i(\eta_{\nu\rho}\gamma_\mu - \eta_{\mu\rho}\gamma_\nu) \tag{18.13}$$

The reduction of $S(\Lambda)$ is achieved with the same γ_F constructed in Section 17.2 when discussing D_l spinors. We find:

$$\begin{aligned} &\{\gamma_\mu^{(0)}, \gamma_F\} = 0 \Rightarrow \{\gamma_\mu, \gamma_F\} = 0 \Rightarrow \\ &[M_{\mu\nu}, \gamma_F] = 0 \Rightarrow \\ &\quad S(\Lambda)\gamma_F = \gamma_F S(\Lambda). \end{aligned} \tag{18.14}$$

Denote the eigenspaces of γ_F for eigenvalues ± 1 as $\mathcal{V}_\pm$: these are the $\mathcal{V}_2$ and $\mathcal{V}_1$ of Section 17.2, Eq.(17.19). Then $S(\Lambda)$ restricted to these 2^{l-1}-dimensional subspaces gives us two irreducible spinor representations $S_\pm(\Lambda)$ of $\mathrm{SO}(p,q)$:

$$\begin{aligned} \mathcal{V} &= \mathcal{V}_+ \oplus \mathcal{V}_-, \\ S(\Lambda) &= \begin{pmatrix} S_+(\Lambda) & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & S_-(\Lambda) \end{pmatrix} \end{aligned} \tag{18.15}$$

18.3 Representations Related to $S(\Lambda)$

One can pass from the Dirac spinor representation of $\mathrm{SO}(p,q)$ to its adjoint, contragredient or conjugate, as defined in Section 8.5. The representation matrices, and their generators, behave as follows:

$$\begin{aligned} S(\Lambda), M_{\mu\nu} &\longrightarrow \text{adjoint: } (S(\Lambda)^\dagger)^{-1}, M_{\mu\nu}^\dagger; \\ &\longrightarrow \text{contragredient: } (S(\Lambda)^T)^{-1}, -M_{\mu\nu}^T; \\ &\longrightarrow \text{conjugate: } S(\Lambda)^*, -M_{\mu\nu}^* \end{aligned} \tag{18.16}$$

To relate each of these to $S(\Lambda)$, we need matrices relating γ_μ to $\gamma_\mu^\dagger, \gamma_\mu^T$ and γ_μ^*. These are the generalisations of the familiar A, B, C matrices of the Dirac equation.

Define the matrix A by

$$A = \gamma_{p+1}\gamma_{p+2}\cdots\gamma_{p+q} = i^q\gamma_{p+1}^{(0)}\cdots\gamma_{p+q}^{(0)} \tag{18.17}$$

It is the product of all the "time-like" gammas. Such a matrix was not needed in the D_l analysis. It obeys

$$\gamma_\mu^\dagger = (-1)^q A\gamma_\mu A^{-1} \tag{18.18}$$

Since the γ_μ are irreducible, such an A is unique upto a factor, and we choose the specific one in Eq.(18.17). To pass to the transpose, we can use C constructed in Section 17.3; it obeys Eq.(17.24) which we repeat here as

$$\gamma_\mu^T = (-1)^l C\gamma_\mu C^{-1} \tag{18.19}$$

Again, the irreducibility of γ_μ means such a C is unique upto a factor. Combining Eqs.(18.18),(18.19) with the unitarity of C we get the behaviour under complex conjugation:

$$\gamma_\mu^* = (-1)^{q+l} CA\gamma_\mu(CA)^{-1} \tag{18.20}$$

The important algebraic properties of A and C are:

$$\begin{aligned} &C^T = C^\dagger = C^{-1} = (-1)^{l(l+1)/2}C, \\ &A^\dagger = A^{-1} = (-1)^{q(q+1)/2}A; \qquad \text{(a)} \\ &A^T = (-1)^{lq+(1/2)q(q-1)}CAC^{-1}, \\ &A^* = (-1)^{q(l+1)}CAC^{-1}. \qquad \text{(b)} \end{aligned} \tag{18.21}$$

Thanks to the relations (18.18),(18.19) and (18.20), we see that the Dirac spinor representation $S(\Lambda)$ goes into itself under each of the three operations (18.16): it is self adjoint, self contragredient and self conjugate. These are achieved thus:

$$\begin{aligned} M_{\mu\nu}^\dagger = AM_{\mu\nu}A^{-1} &: \ (S(\Lambda)^\dagger)^{-1} = AS(\Lambda)A^{-1}; \\ -M_{\mu\nu}^T = CM_{\mu\nu}C^{-1} &: \ (S(\Lambda)^T)^{-1} = CS(\Lambda)C^{-1}; \\ -M_{\mu\nu}^* = CAM_{\mu\nu}(CA)^{-1} &: \ S(\Lambda)^* = CAS(\Lambda)(CA)^{-1} \end{aligned} \tag{18.22}$$

In these relations since $S(\Lambda)$ is reducible we could have replaced A and C by $Af(\gamma_F)$ and $Cg(\gamma_F)$, when $f(\gamma_F)$ and $g(\gamma_F)$ are nonsingular.

18.4 Behaviour of the Irreducible Spinor Representations $S_{\pm}(\Lambda)$

The consequences of Eqs.(18.22) for the irreducible representations $S_{\pm}(\Lambda)$ contained in $S(\Lambda)$ depend on the behaviours of A and C with respect to γ_F, and so ultimately on the parities of l and q. We find that

$$\begin{aligned} C\gamma_F &= (-1)^l \gamma_F C, \\ A\gamma_F &= (-1)^q \gamma_F A, \\ CA\gamma_F &= (-1)^{q+l} \gamma_F CA \end{aligned} \tag{18.23}$$

It is useful to recall that if we have a Dirac spinor ψ belonging to the representation $S(\Lambda)$, then

$$\begin{aligned} \psi \to \psi' = S(\Lambda)\psi \Rightarrow \\ \psi^\dagger \to \psi'^\dagger = \psi^\dagger A S(\Lambda)^{-1} A^{-1} \Rightarrow \\ \bar{\psi} \to \bar{\psi}' = \bar{\psi} S(\Lambda)^{-1}, \bar{\psi} = \psi^\dagger A \Rightarrow \\ \bar{\psi}\psi = \text{invariant} \end{aligned} \tag{18.24}$$

On putting together the reduction (18.15) of $S(\Lambda)$ and the properties (18.23), we get the following pattern of results:

l	q	*Equivalent to* S_+	*Equivalent to* S_-	*Remark on* $\bar{\psi}\psi$
even	even	$(S_+^\dagger)^{-1}, (S_+^T)^{-1}, S_+^*$	$(S_-^\dagger)^{-1}(S_-^T)^{-1}, S_-^*$	No mixing of chiralities
even	odd	$(S_+^T)^{-1}, (S_-^\dagger)^{-1}, S_-^*$	$(S_-^T)^{-1}, (S_+^\dagger)^{-1}, S_+^*$	Mixes chiralities
odd	even	$(S_+^\dagger)^{-1}, (S_-^T)^{-1}, S_-^*$	$(S_-^\dagger)^{-1}, (S_+^T)^{-1}, S_+^*$	No mixing of chiralities
odd	odd	$S_+^*, (S_-^\dagger)^{-1}, (S_-^T)^{-1}$	$S_-^*, (S_+^\dagger)^{-1}, (S_+^T)^{-1}$	Mixes chiralities

One can see that the "Dirac mass term" $\bar{\psi}\psi$ mixes chiralities if q, the number of time like dimensions, is odd, and not if it is even. The above list of properties is relevant in discussing Weyl and Majorana spinors.

18.5 Spinor Representations of SO(p,q) for $p+q=2l+1$

The Dirac algebra reads

$$\{\gamma_\mu,\gamma_\nu\}=2\eta_{\mu\nu},\quad \mu,\nu=1,2,\ldots,2l,2l+1 \tag{18.25}$$

On the space $\mathcal{V}$ of dimension 2^l, we have two irreducible inequivalent representations possible:

$$\begin{aligned}
&\text{(i) } \gamma_j=\gamma_j^{(0)};\gamma_r=i\gamma_r^{(0)},r=p+1,\ldots,p+q-1=2l;\\
&\qquad \gamma_{2l+1}=i\gamma_F\\
&\text{(ii) } \gamma_j=\gamma_j^{(0)};\gamma_r=i\gamma_r^{(0)},r=p+1,\ldots,p+q-1=2l;\\
&\qquad \gamma_{2l+1}=-i\gamma_F
\end{aligned} \tag{18.26}$$

But, as it happened in the case of B_l, they will lead to equivalent spinor representations and generators, so we stick to representation (i) above in the sequel.

The Dirac spinor representation and its generators are now irreducible on $\mathcal{V}$:

$$\begin{aligned}
\Lambda\in \mathrm{SO}(p,q)\cdot \gamma'_\mu&=\Lambda^\nu{}_\mu\gamma_\nu=S(\Lambda)\gamma_\mu S(\Lambda)^{-1},\\
S(\Lambda')S(\Lambda)&=\pm S(\Lambda'\Lambda),\\
S(\Lambda)&\sim\exp\left(-\frac{i}{2}\omega^{\mu\nu}M_{\mu\nu}\right),\\
M_{\mu\nu}&=\frac{i}{4}[\gamma_\mu,\gamma_\nu]
\end{aligned} \tag{18.27}$$

The hermiticity and commutation relations (18.3), (18.4) are again satisfied.

Irreducibility of $S(\Lambda)$ immediately tells us it must be again self adjoint, self contragredient and self conjugate. For these purposes, the same C matrix can be used as previously, but the A matrix is different:

$$\begin{aligned}
A&=\gamma_{p+1}\gamma_{p+2}\cdots\gamma_{p+q}\\
&=i^q\gamma_{p+1}^{(0)}\cdots\gamma_{2l}^{(0)}\gamma_F\\
&=i\times(A\text{-matrix for }\mathrm{SO}(p,q-1))\times\gamma_F
\end{aligned} \tag{18.28}$$

One has again the adjoint and other properties

$$\begin{aligned}
\gamma_\mu^\dagger&=(-1)^qA\gamma_\mu A^{-1},\\
\gamma_\mu^T&=(-1)^lC\gamma_\mu C^{-1},\\
\gamma_\mu^*&=(-1)^{q+l}CA\gamma_\mu(CA)^{-1},\quad \mu=1,2,\ldots,2l+1
\end{aligned} \tag{18.29}$$

Moreover, even with the new definition of A, *all of Eqs.(18.21), (18.22) continue to be valid with no changes at all*, so we do not repeat them. The only difference is that since $M_{\mu\nu}$ and $S(\Lambda)$ are now irreducible, there is now no freedom to attach non-singular functions of γ_F to A and C in Eq.(18.22).

18.6 Dirac, Weyl and Majorana Spinors for SO(p,q)

For both $p+q=2l=$ even, and $p+q=2l+1=$ odd, we work in the same space $\mathcal{V}$ of dimension 2^l. In the even case, $S(\Lambda)$ reduces to the irreducible parts $S_\pm(\Lambda)$ (corresponding to $\gamma_F=\pm 1$); in the odd case, it is irreducible.

For any $p+q$: a *Dirac* spinor is any element ψ in the linear space $\mathcal{V}$, subject to the transformations $S(\Lambda)$ as in Eq.(18.24). Weyl and Majorana spinors are Dirac spinors obeying additional conditions.

Weyl Spinors

These are defined only when $p+q=2l$. A *Weyl* spinor is a Dirac spinor ψ obeying the Weyl condition:

$$\gamma_F\psi=\epsilon_\omega\psi, \epsilon_\omega=\pm 1. \tag{18.30}$$

Depending on ϵ_ω, we get righthanded (positive chirality) or lefthanded (negative chirality) spinors:

$$\begin{aligned}&\epsilon_\omega=+1:\ \psi\in\mathcal{V}_+, \psi'=S_+(\Lambda)\psi:\ \text{right handed}\\ &\epsilon_\omega=-1:\ \psi\in\mathcal{V}_-, \psi'=S_-(\Lambda)\psi:\ \text{left handed}\end{aligned} \tag{18.31}$$

If $p+q=2l+1$, we do not define Weyl spinors at all.

Majorana Spinors

These can be considered for any $p+q$, and are Dirac spinors obeying a reality condition. We know that for any $p+q$, we have

$$S(\Lambda)^*=CAS(\Lambda)(CA)^{-1} \tag{18.32}$$

so for any Dirac spinor ψ,

$$\begin{aligned}\psi'=S(\Lambda)\psi\Rightarrow\\ (CA)^{-1}\psi'^*=S(\Lambda)(CA)^{-1}\psi^*\end{aligned} \tag{18.33}$$

That is, a Dirac ψ and $(CA)^{-1}\psi^*$ transform in the same way. We have a Majorana spinor if these two are essentially the same. Three possible situations can arise, which we look at in sequence. First we gather information regarding two important phases ξ,η connected with CA. For any p,q we have under transposition:

$$\begin{aligned}(CA)^T&=\xi CA,\\ \xi&=(-1)^{\frac{1}{2}l(l+1)+\frac{1}{2}q(q-1)+lq}\end{aligned} \tag{18.34}$$

Here we have used Eq.(18.21(b)), and have left the dependence of ξ on l and q implicit. [Remember also that the definition of A depends on whether $p+q$ is even or odd, but that since Eqs.(18.21) are uniformly valid, so is Eq.(18.34) above]. On the other hand, all the matrices C, A, CA are unitary, so we also have

$$(CA)^*(CA) = \xi \tag{18.35}$$

With respect to γ_F, we have the property (18.23) *relevant only if* $p + q = 2l =$ *even*, and written now as

$$CA\gamma_F = \eta\gamma_F CA, \eta = (-1)^{l+q} \tag{18.36}$$

So, the phase ξ is defined for all p and q, whether $p+q = 2l$ or $2l+1$; while the phase η is defined only when $p+q = 2l$.

The three situations are as follows:

(i) Suppose $p+q = 2l+1$, so $S(\Lambda)$ is irreducible. A Majorana spinor is a Dirac spinor which for some complex number α obeys

$$\begin{aligned} &(CA)^{-1}\psi^* = \alpha\psi \\ \text{i.e.,} \qquad &\psi^* = \alpha CA\psi \end{aligned} \tag{18.37}$$

Taking the complex conjugate of this condition, using it over again, and then Eq.(18.35), one finds the restriction

$$|\alpha|^2\xi = 1 \tag{18.38}$$

Thus, α must be a pure phase, which can be absorbed into ψ. A Majorana spinor can therefore exist in an odd number of dimensions if and only if the phase $\xi = 1$. The sufficiency is shown by the following argument: if $\xi = 1$, then the matrix CA is both unitary and symmetric, so by the argument of section 10 it possesses a complete orthonormal set of real eigenvectors. Thus nonvanishing ψ obeying the Majorana condition (18.37) do exist.

(ii) Suppose $p+q = 2l$, and we ask if there are spinors having both Weyl and Majorana properties. Now $S(\Lambda)$ is reducible, so we must contend with the fact that ψ^* transforms in the same way as $f(\gamma_F)CA\psi$ for any function $f(\,\cdot\,)$. Let us ask for the necessary and sufficient conditions for existence of Majorana–Weyl spinors. The Weyl condition says.

$$\begin{aligned} &\gamma_F\psi = \pm\psi, \\ \text{i.e.,} \qquad &\psi = \begin{pmatrix} \varphi \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} 0 \\ \chi \end{pmatrix} \end{aligned} \tag{18.39}$$

Given that ψ is of one of these two forms, the Majorana condition says

$$\psi^* = \alpha CA\psi \tag{18.40}$$

where there is no need to include the factor $f(\gamma_F)$. Unless CA is block diagonal, ψ would vanish, so a necessary condition is $\eta = 1$. Given this, it follows that

$\xi = 1$ is also a necessary condition. But these are also sufficient! For, if $\xi = \eta = 1$, we have

$$CA = \begin{pmatrix} K_1 & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & K_2 \end{pmatrix} \tag{18.41}$$

with both K_1 and K_2 being unitary and symmetric. Then each of K_1 and K_2 has a complete orthonormal set of *real* eigenvectors, and we can have Majorana–Weyl spinors with $\epsilon_\omega = +1$ or -1.

Thus, $\xi = \eta = 1$ is the necessary and sufficient condition for Majorana–Weyl spinors to exist. In this case, plain Majorana spinors not having the Weyl property are just obtained by putting together Majorana–Weyl spinors with both ϵ_ω values, and the general Majorana condition is

$$\psi^* = f(\gamma_F)CA\psi \tag{18.42}$$

(iii) Suppose $p + q = 2l$, but Majorana–Weyl spinors do not exist. Then either $\xi = -1, \eta = 1$ or $\xi = 1, \eta = -1$ or $\xi = \eta = -1$. In each case let us see if Majorana spinors exist.

If $\xi = -1, \eta = 1$, and we write CA in the block diagonal form (18.41), then both K_1 and K_2 are unitary antisymmetric:

$$K_1^* K_1 = K_2^* K_2 = -1 \tag{18.43}$$

The most general Majorana condition on ψ is

$$\begin{aligned} &\psi^* = f(\gamma_F)CA\psi, \\ \text{i.e.,} \quad &\begin{pmatrix} \varphi^* \\ \chi^* \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix} \begin{pmatrix} \varphi \\ \chi \end{pmatrix}, \end{aligned} \tag{18.44}$$

where α, β are complex numbers. But the fact that $\xi = -1$ forces both φ and χ to vanish, so there are no Majorana spinors at all.

If $\eta = -1, \xi = \pm 1$, then CA is block-off-diagonal:

$$\begin{aligned} CA &= \begin{pmatrix} 0 & K \\ \xi K^T & 0 \end{pmatrix}, \\ K^\dagger K &= 1 \end{aligned} \tag{18.45}$$

The general Majorana condition on ψ is

$$\begin{aligned} &\psi^* = f(\gamma_F)CA\psi, \\ \text{i.e.,} \quad &\begin{pmatrix} \varphi^* \\ \chi^* \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} 0 & K \\ \xi K^T & 0 \end{pmatrix} \begin{pmatrix} \varphi \\ \chi \end{pmatrix}, \\ \text{i.e.,} \quad &\varphi^* = \alpha K \chi, \chi^* = \beta \xi K^T \varphi \\ \text{i.e.,} \quad &\varphi = \alpha^* K^* \beta \xi K^T \varphi \\ \text{i.e.,} \quad &\alpha^* \beta = \xi \end{aligned} \tag{18.46}$$

Thus, both α and β must be non-zero, and similarly both φ and χ must be non-zero, and such Majorana spinors definitely exist.

Collecting all the above results, we have the following picture telling us when each kind of spinor can be found:

	Weyl	**Majorana**	**Majorana–Weyl**
$p+q=2l+1$	$\times$	$\xi=1$	$\times$
$p+q=2l$	Always, $\epsilon_\omega=\pm 1$	$\xi=\eta=1$; $\xi=\pm 1, \eta=-1$	$\xi=\eta=1$

All these results are based on group representation structures alone. We have also assumed that each ψ is a "vector" in $\mathcal{V}$ with complex numbers for its components. If they are Grassmann variables, or if we ask for the Weyl or Majorana or combined property to be maintained in the course of time as controlled by some field equations, other new conditions may arise.

As an elementary application let us ask in the case of the groups $SO(p,1)$ when we can have Majorana–Weyl spinors. Now $q=1$ and so p must necessarily be odd so that $p+1=2l$ can be even. The necessary and sufficient conditions are

$$\xi=(-1)^{\frac{1}{2}l(l+1)+l}=1,$$
$$\eta=(-1)^{l+1}=1$$

This limits l to the values $l=5,9,13,\ldots$ (omitting the somewhat trivial case $l=1$), so we have Majorana–Weyl spinors only for the groups SO(9, 1), SO(17, 1), SO(25, 1)... in the family of Lorentz groups $SO(p,1)$.

Exercises for Chapters 17 and 18

1. Show that an alternative representation for the gamma matrices of Eq.(17.7) is
$$\gamma_{ar}=\sigma_3^{(1)}\sigma_3^{(2)}\cdots\sigma_3^{(a-1)}\sigma_r^{(a)}.$$

2. For the construction of the spinor representations of $\mathcal{D}_l=SO(2l)$ in Sections 17.1, 17.2, reconcile the irreducibility of the γ_A with the reducibility of the generators M_{AB}.

3. For the group SO(6) which is locally isomorphic to SU(4), trace the connection of the two spinor UIR's $\Delta^{(1)}$ and $\Delta^{(2)}$ to the defining representation of SU(4) and its complex conjugate.

4. For the case $l=3, \mathcal{D}_l=SO(6)$, supply proofs of the results stated in the table at the end of Section 17.4.

5. Verify all the stated relations in Eqs.(18.18), (18.19), (18.20), (18.21), (18.22). Similarly for Eqs.(18.29).

6. Check that the 4 component spinor ψ in the familiar Dirac wave equation is the direct sum of two irreducible two-component spinors of SO(3, 1) transforming as complex conjugates of one another.

Bibliography

[1] R. Brauer and H. Weyl: 'Spinors in n dimensions', *Am. J. Math.* **57**, 425–449 (1935).

[2] G. Racah: 'Group Theory and Spectroscopy', *Princeton lectures*, **1951**, reprinted in *Ergebn. der. Exakten Naturwiss*, **37**, 28–84 (1965).

[3] W. Pauli: 'Continuous Groups in Quantum Mechanics,' *CERN Lectures*, **1956**; reprinted in *Ergebn. der. Exakten Naturwiss.* **37**, 85–104 (1965).

[4] L. Pontrjagin: *Topological Groups*, Princeton University Press, 1958.

[5] R.E. Behrends; J. Dreitlein, C. Fronsdal and B.W. Lee, "Simple Groups and Strong Interaction Symmetries", *Revs. Mod. Phys.*, **34**, 1–40, 1962.

[6] M. Hamermesh: *Group Theory and its Applications to Physical Problems*, Addison Wesley, 1962.

[7] A. Salam: "The Formalism of Lie Groups", *Trieste Lectures*, 1963.

[8] H. Boerner: *Representations of Groups with special consideration for the needs of Modern Physics*, North Holland, 1970.

[9] E.C.G. Sudarshan and N. Mukunda: *Classical Dynamics — A Modern Perspective*, Wiley, New York, 1974.

[10] R. Gilmore: *Lie Groups, Lie Algebras and Some of their Applications*, Wiley, New York, 1974.

[11] B.G. Wybourne: *Classical Groups for Physicists*, Wiley, New York, 1974.

[12] R. Slansky: 'Group Theory for Unified Model Building', *Phys. Rep.*, **79C**, 1, 1981.

[13] P. Langacker: 'Grand Unified Theories and Proton Decay', *Phys. Rep.*, **72C**, 185, 1981.

[14] J. Wess: 'Introduction to Gauge Theory', *Les Houches Lectures*, 1981.

[15] H. Georgi: *Lie Algebras in Particle Physics*, Benjamin/Cummings, 1982.

[16] F. Wilczek and A. Zee: 'Families from Spinors', *Phys. Rev.*, **D25**, 553, 1982.

[17] P. van Nieuwenhuizen: "Simple supergravity and the Kaluza-Klein Program," *Les Houches Lectures*, 1983.

[18] A.P. Balachandran and C.G. Trahern: *Lectures on Group Theory for Physicists*, Bibliopolis, Napoli, 1984.

[19] A.O. Barut and R. Raczka: "Theory of Group Representations and Applications," World Scientific, Singapore, 1986.

Index

Lectures on **Advanced Mathematical Methods** for **Physicists**

This book presents a survey of Topology and Differential Geometry and also, Lie Groups and Algebras, and their Representations. The first topic is indispensable to students of gravitation and related areas of modern physics (including string theory), while the second has applications in gauge theory and particle physics, integrable systems and nuclear physics.

Part I provides a simple introduction to basic topology, followed by a survey of homotopy. Calculus of differentiable manifolds is then developed, and a Riemannian metric is introduced along with the key concepts of connections and curvature. The final chapters lay out the basic notions of simplicial homology and de Rham cohomology as well as fibre bundles, particularly tangent and cotangent bundles.

Part II starts with a review of group theory, followed by the basics of representation theory. A thorough description of Lie groups and algebras is presented with their structure constants and linear representations. Root systems and their classifications are detailed, and this section of the book concludes with the description of representations of simple Lie algebras, emphasizing spinor representations of orthogonal and pseudo-orthogonal groups.

The style of presentation is succinct and precise. Involved mathematical proofs that are not of primary importance to physics student are omitted. The book aims to provide the reader access to a wide variety of sources in the current literature, in addition to being a textbook of advanced mathematical methods for physicists.

World Scientific
www.worldscientific.com
7693 hc

ISBN-13 978-981-4299-73-2
ISBN-10 981-4299-73-1

9 789814 299732